STIC Scientific and Technical Information Center	Dialog®

Core FT1:

- Business & Industry , File 9 (1994 - present)
- ABI/INFORM®, File 15 (1971 - present)
- Gale Group PROMT®, File 16 (1990 - present)
- Gale Group PROMT®, File 160 (1972-1989)
- Gale Group Trade & Industry Database , File 148 (1976 - present)
- Gale Group Computer Database , File 275 (full-text 1/1988 - present)
- Business Wire, File 610 (Mar 1999 - present)
- Business Wire, File 810 (1986 - February 1999)

Core FT2:

- Dialog Global Reporter, File 20 (May 1997 - present)
- The McGraw-Hill Companies Publications Online, File 624 (1985 - present)
- Gale Group New Product Announcements/Plus® (NPA/Plus, File 621 (1985 - present)
- Gale Group Newsletter Database , File 636 (1988 - present)
- PR Newswire, File 613 (May 1999 - present)
- San Jose Mercury News, File 634 (Jun 1985 - present)
- PR Newswire, File 813 (May 1987 - May 1999)

Sub35FT:

- McClatchy-Tribune Information Service, File 608 (Jan 1989 - present)
- American Banker Financial Publications, File 625 (1981 - June 2008)
- Banking Information Source, File 268 full-text (1994 - present)
- Bond Buyer Full Text, File 626 (November 1981 - April 2008)
- DIALOG Finance & Banking Newsletters, File 267 (1996 - present)

Set#	Query
L1	RANDOM$2 ADJ3 SLECT$4
L2	historical adj interval
L3	(RANDOM$2 ADJ3 SeLECT$4) same (historical$2 adj interval$2)

L4	(RANDOM$2 with SeLECT$4) and (historical$2 with interval$2) and ((financ$4 portfolio) with (analysis value valuation))
L5	(random$2 with select$4) and (historical$2 with interval$2) and ((financ$4 portfolio) same (analysis value valuation))
L6	(RANDOM$2 and historical$2 and interval$2) and ((financ$4 portfolio) and (analysis value valuation))

36/9/1 (Item 1 from file: 268)
00383985 53790205
Benchmarking manager performance within the private real estate investment industry
Geltner, David
Real Estate Finance , v 17 , n 1 , p 23-34 , Spring 2000 **Document Type:** Periodical; Feature **ISSN:** 0748-318X **Journal Code:** RFN **Language:** English **Record Type:** Fulltext
ARTICLE REFERENCE NUMBER: RFN-2019-5
Word Count: 8,758

Abstract:

What makes a good benchmark for evaluating institutional investment manager performance in the private real estate investment industry? Is the NCREIF index a good benchmark? How can you appropriately use the NCREIF index in such a role? This article suggests that the criteria used to judge and design good benchmark in the public securities industry, do not completely carry on over to the private real estate business. Some criticisms of the NCREIF index as a benchmark come from the public securities perspective, and in the real world of private real estate tend to have the characteristics of attacks against a straw man, or attacks against benchmarking itself, or strivings for a will-o'-the-wisp.

Text:

SUDDENLY, LAST DECADE...

Once there was a time, a little over a decade ago, when the NCREIF index was the real estate investment manager's friend. It was hailed as representing the institutional real estate investment universe, representing virtually all the properties held for tax-exempt investors by virtually all the major investment management firms. And it tracked those properties' investment performance in such a way that this performance could be compared to that of stocks and bonds, and used in modern **portfolio** theory (MPT) models of asset allocation for pension fund portfolios. It showed only a smooth line going in the northeastward direction on the chart. The MPT models said everybody should invest 150% of all their wealth in real estate. And everybody was happy And nobody asked very pointed questions like, how did you do compared to the Index?

And then suddenly last decade... something happened. Well, a lot happened. First, real estate crashed and there were a lot of unhappy campers. Then the client side of the investment industry got a lot more serious and sophisticated and started asking questions about manager performance, not just in absolute terms, but relative to the "industry," that is, relative to the "peer universe," in other words, relative to the

NCREIF index. People started worrying about alignment of interests. The easy times were over, not only for investment managers, but also for the NCREIF index, because people started using the index as a "benchmark. "

Suddenly, the NCREIF index really mattered. It could affect your pocketbook, and not just in positive ways. Vast numbers of critical eyes were turned on the index, and they found it wanting. It's "appraisal-based," they murmured. It's "lagged and smoothed," they muttered. It's "not passive," they complained. It's "not investable," they cried. It's "not representative," they shouted. It's "ambiguous," they shuddered.

These complaints are all true. And they are all valid points. But let's not panic. Let's look at things with a clear eye, and go back to the basics: What is the role and purpose of benchmarking? What are the strengths and weaknesses of the NCREIF index in serving this role and purpose? Let's not just assume that what makes sense in the public securities investment industry must necessarily be exactly copied in the private real estate investment industry for the latter industry to serve the needs of institutional investors.

It is important that we go through this exercise very carefully, because if the NCREIF index cannot be used as a benchmark, if the above complaints are fatal to such a role for the NCREIF index, then it is reasonable to ask whether it is possible to have any workable benchmark in the private real estate investment industry. And if it is not possible to benchmark private real estate investment performance, then the entire industry may be in jeopardy We may be in for a paradigm shift in which real estate will no longer be viewed as a core asset class in the institutional **portfolio**. The stakes are high....

THE BASIC IDEA

Performance evaluation refers to the process of judging how well an investment manager has performed. While this process broadly involves considerations that cannot be well quantified, there are aspects of the investment management job that can be quantified, and it is in the quantitative aspect of evaluation that benchmarking is most relevant.

A typical performance evaluation of an investment manager or an investment fund is presented in the Exhibit, which depicts the evaluation of two managers, "A" and "B." After a period of time, typically three to five years, the average return achieved over that period is computed for each manager.1 The returns achieved over that same period of time are also computed for each manager in a peer universe of all the competing managers (or funds) who also specialize in the same type of investments.

In the Exhibit, the manager's performance is indicated by the cross. The vertical line indicates the fifth and ninety-fifth percentile range of the manager's peer universe. The box indicates the middle two quartiles, and the horizontal line inside the box indicates the median manager's performance within the peer universe. The triangle indicates the return recorded by a passive index that is broadly representative of the type of assets the manager specializes in. (A \passive" index is one whose allocations across assets remain relatively constant.)

In the example shown in the Exhibit, Manager A did quite well, performing in the upper quartile of the peer universe, and substantially beating the relevant passive index. Manager A will likely be rehired and may have earned an incentive fee. It doesn't look so good for Manager B. It is likely the investor will fire Manager B or reduce the allocation in the future, probably hiring one of Manger B's competitors instead. (Note that Manager A probably is not a competitor to Manager B, as they are in two different asset classes or management styles.2)

BENCHMARKS IN THE PRIVATE REAL ESTATE ASSET CLASS

In the Exhibit, the manager's performance is compared both with respect to the median of her peer universe, and with respect to a passive index broadly representative of her asset class. For example, in the public equity investment business, if the manager specializes in large-capitalization stocks, then the peer universe would consist of all

the major mutual funds with a similar specialty, and the passive index most widely used to benchmark large-stock performance is the Standard & Poor's Composite 500 index (the S&P 500).

In the example shown in the Exhibit, Manager A has beaten both the passive index and her average competitor. Manager B has been beaten by both of these reference points. More generally, however, there is potential ambiguity in this process. A manager might, for example, beat an average competitor but fall short of the relevant passive index, or vice versa.

So, which type of reference point is better as a benchmark for judging manager performance? Should it be some central tendency of the manager's peer universe (say, the mean or median), or should it be a passive index? In the securities investment industry, both types of benchmarks are used, but there is a general conception that passive indexes are a bit better, in principle. This is primarily for two related reasons:

EXHIBIT

A Graphical Depiction of Typical Investment Manager Performance Evaluation

In the public securities markets, passive indexes are essentially replicable by both the manager and the client.

Therefore, the client can presumably invest in the passive index as an alternative to hiring an active manager.

In contrast to stock market investments, in the private real estate asset class it is not really possible to have a truly passive index. Fundamentally, this is because private real estate investment managers have to be responsible for the operational management of the properties they own on behalf of their clients. Furthermore, private real estate investment necessarily involves long holding periods, because transaction costs are high. Therefore, much of the overall investment performance is attributable to how well properties are managed while they are held. This is "active management," of necessity.3

If it is not possible to have a passive index in the private real estate asset class, then to claim that the NCREIF index cannot be used as a benchmark because it is not a passive index is more than just a criticism of the NCREIF index. It amounts to a claim that it is impossible to have any benchmark in the private real estate investment industry.

But is it really necessary for a benchmark to be a passive index? A good argument can be made that the major reasons why passive indexes make better benchmarks do not apply within the private real estate asset class. First, because of the long holding periods, the segment allocation weights in the relevant private real estate peer universe are likely to be much more stable over time than those of the typical actively managed **portfolio** in the securities markets. This makes the peer universe at least somewhat replicable, in terms of segment allocation.

Second, even if we had a static **portfolio** index with constant segment weightings in private real estate, such an index would not present the investor with an alternative to hiring an active manager. For one thing, in the private property market, whole assets are traded. Any specific assets in any static **portfolio** index would already be owned by someone else. Neither the investor directly, nor any manager, could exactly replicate the passive index in the strict sense of holding the same assets. For another thing, passive investors cannot generally invest directly, by themselves, in a **portfolio** of properties, as an alternative to hiring an active manager, because of the specialized expertise that is required for direct investment in the private property market (including the need for operational management expertise).

Instead of passive indexes like the S&P 500, the only type of periodic return indexes available to serve as benchmarks in the private real estate asset class are peer group indexes. These are indexes, like the NCREIF index in the U.S. and the IPD index in Great Britain, that essentially reflect the property-level performance of all (or most) of the private real estate investment managers in the country 4 As such, they correspond conceptually to the central tendency of a given manager's

relevant peer universes In essence, peer universe indexes are the only type of benchmarks likely to be available in the foreseeable future in the private real estate investment industry.6

MATCHING EVALUATION, RESPONSIBILITY, AND AUTHORITY

In selecting (or constructing) a benchmark for use in manager performance evaluation, it is important to adhere to the basic management principle of equating responsibility with authority In general, an investment manager should not be held responsible for, or evaluated on the basis of, factors that are beyond his decision authority or control, especially if the client has explicitly denied the manager discretion over such factors.

This issue arises commonly in the area of segment allocation. Asset market segments in private real estate typically correspond to space market segments, characterized by property type and geographic location. Due to the need for local expertise, it is not uncommon for real estate investment managers to be specialized within one or a few market segments. For example, a manager may be hired by a client for the specific purpose of placing capital into the apartment property segment, perhaps within a proscribed geographic region. It would generally be unfair or misleading to benchmark such a manager's performance to an index that included other types of property besides apartments, and/or other geographic regions beyond the manager's purview.

On the other hand, if a manager is given discretion to choose allocational weights across more than one property market segment, then his benchmark should include all the segments in the manager's potential choice set. In this case, the question is what segment weightings to use in the manager's benchmark.

In the private real estate asset class, there is probably no good general answer to this question. The appropriate benchmark will depend on the client's overall investment strategy and the client's particular objectives for the given manager. However, a couple of general principles can be enumerated:

The segment weights in the manager's benchmark should be mutually agreed upon by the client and the manager in advance, that is, at the outset of the management contract period. (Midcourse changes should be made only with the consent of both parties.)

It should be possible for the manager to at least approximately replicate his benchmark's segment allocation weights, if he wants to. In other words, if the manager does not wish to place a "bet" against the segment allocation in his benchmark, he should not have to.

Both of these principles suggest that the benchmark segment weights should be constant, or at least nearly so, over the period of the manager's contract.' Yet the segment weights in a peer universe-based index such as the NCREIF are not constant over time. However, for the reasons previously noted, segment allocations change only gradually in the typical private property return index. Furthermore, a manager's custom benchmark can be defined using property-level (i.e., within-segment) returns from the peer universe index, rebalanced to reflect constant segment weights agreed upon by the manager and client, or in fact whatever segment weights are agreed upon by the client and the manager.8

In the absence of a compelling reason otherwise, the simplest segment weights to use in the manager's custom benchmark would be equal weights across all the segments within the manager's purview Alternatively in the stock market, the capital asset pricing model (CAPM) is often invoked as a basis to suggest that the segment weights in a manager's benchmark index should reflect the "market **portfolio**." In real estate, this would reflect the market **value**-based segment weights of all commercial property in the country (or at least, in the case of a domestic institutional "core" manager, for example, the market would consist of all "institutional-quality" commercial properties in the country, a somewhat vaguely defined term).

However, the CAPM does not work very well as a descriptor of reality

within the private real estate asset class.9 There may therefore be less theoretical rationale for the use of market-based weights in a manager's benchmark in private real estate investment, as compared to stock market investments.10

On the other hand, a client and/or manager may feel that market weights make at least as good a benchmark as any others. In particular, market weights approximately reflect the relative abilities of different segments to absorb new investment without price distortion. In other words, market weights may approximate the relative ease for the manager in finding assets for sale in the different segments. In this sense, market weights may approximate something like "passive index" segment weightings.11

THE PROBLEM OF STATISTICAL SIGNIFICANCE

When people seek to quantitatively evaluate investment manager performance as compared to a benchmark, the underlying purpose of such evaluation is usually perceived to be the identification of "superior" versus "inferior" investment managers in terms of their investment abilities. Investors want to place their capital with superior managers, and avoid using inferior managers. Rigorously speaking, what is meant by "superior" investment ability is the ability to consistently beat an appropriate benchmark, as opposed to a **random** outcome in which a manager happens to beat the benchmark over a given period of **historical** time. Similarly, "inferior" ability is the opposite, the tendency to get consistently beaten by the appropriate benchmark, as distinguished from a **random** streak of "bad luck."

This perception of the purpose of performance evaluation raises a fundamental conceptual problem in practice. This problem exists whether evaluation is done in the context of the public securities market or the private real estate market. The problem is a lack of "statistical significance" in ex post performance differentials. What does this mean?

A basic characteristic of risky assets is that their realized returns vary **randomly** around their rational ex ante expectations. This randomness in ex post returns is simply a reflection of the risk in the assets from an investment perspective. Performance evaluation must be based on ex post returns, the only kind of returns that can be objectively quantified and that reflect the actual realized performance of the manager. The existence of randomness in these returns means that performance evaluation is an exercise in statistical inference. Based on a sample of time (the history over which the performance evaluation is being conducted), we are trying to infer what is the central tendency of the manager's differential performance relative to the benchmark over all (or any given **interval** of) time.'Z The greater the randomness in ex post returns, and the shorter the **historical** time sample for which we have differential performance data, the less accurately we can infer the central tendency from the available empirical evidence.

Consider a simple numerical example. Suppose a manager's quarterly holding-period returns are compared against those of an appropriate benchmark over a three-year period. That is, we have twelve observations (or "sample drawings") about the underlying "true" (or "population") difference between Manager i's quarterly return and the benchmark's quarterly return. (We call this true difference E[r

sub i

−r

sub B

].) Suppose we have very favorable circumstances for inferring this difference. In particular, suppose the true difference is constant over time, and we observe the realized returns to both the manager and the benchmark without error. Suppose also that the true quarterly volatility is only 5% in both ri and rB, and the correlation between the manager's returns and the benchmark returns is 90%.13 Then Manager i's average quarterly return would have to exceed the average benchmark return by over 130 basis points (an annualized return difference of over 5/) before we could conclude with statistical significance that the central tendency of

the manager's return exceeds that of the benchmark's return (in other words, that the manager can beat the benchmark as a result of skill rather than a **random** outcome.)14 But it is very rare to find a manager's return beating the benchmark by this much.15

As a result of such considerations, at least one major national real estate investment industry association, in Australia, has formally recommended that a "health warning" should accompany presentation of real estate investment performance results. The suggested wording as of 1998 was:16

Past investment performance is not an adequate test of comparative performance, nor a reliable indicator of the expected absolute level of returns in the future.

IMPLICATIONS OF THE LACK OF STATISTICAL SIGNIFICANCE: "WILL THE REAL PURPOSE OF PERFORMANCE EVALUATION PLEASE STANDUP?"

At first glance, it would seem that the difficulty of making rigorous inferences about managers' investment abilities on the basis of the evidence in their realized performance would render quantitative performance evaluation pointless. Could a simple statistical point really undercut an endeavor that so many people spend so much effort on in the real world? Probably not, for at least a couple of reasons.

Including Non-Quantitative Considerations

Investors rarely base their judgments about manager performance solely on the basis of quantitative **analysis**. Managers generally explain and describe their investment and management "philosophy," their strategy, tactics, and management procedures, in some detail for their clients or potential clients. Investors make judgments in part based on these "stories." If the story makes sense, the investor will factor this non-quantitative information in with the quantitative evidence. The two types of information together may allow a substantially more accurate judgment to be made about a given manager's ability, at least relative to his competitors.

The Ex Ante Role of Performance Evaluation

The critique of performance evaluation from a statistical significance perspective in the preceding section viewed the purpose of such evaluation from a rather narrow, ex post perspective. We suggested that, if performance evaluation is for the purpose of inferring, ex post, which managers are truly "superior" and which are truly "inferior" (based on skill), then it will often be very difficult to achieve this purpose. But this is surely not the only, and perhaps not even the primary purpose of investment manager performance evaluation (even though it may be consciously perceived as such by many practitioners). In fact, performance evaluation serves an ex ante purpose that may be more important than its ex post purpose. To see this, consider the following characterization of the relationship between investors and managers.

Suppose the typical investor hires the typical manager in a three-year contract, at the end of which the manager's performance will be formally evaluated. An incentive fee or bonus based on this retrospective evaluation may (though need not necessarily) be included in the management contract.17 The performance evaluation that occurs at the end of the contract formally (and empirically) applies to the preceding three years, the years during which the manager worked for the client. In this formal ex post usage, the evaluation probably will lack statistical significance. But the mere prospect that the evaluation will be performed has an impact on the manager's incentives and behavior ex ante, that is, prior to the expiration of the contract. By working harder (or "smarter") for the client, the manager increases the likelihood (ex ante) of beating their benchmark.18 This increases the likelihood that the manager will be rehired for a subsequent multiyear period (and it increases the manager's expected earnings from any incentive fee component of the contract based on the formal evaluation). Thus, on an ex ante basis, the use of formal, quantitative performance evaluation helps to align the interests of the manager and the investor, thereby reducing agency costs in the investment

industry.

Ex post, "water is over the dam," and there is often little that can be concluded, rigorously speaking. The "real purpose" (or more important role) of formal, quantitative investment performance evaluation probably lies in the ex ante role just described, particularly in combination with the (often non-quantitative) plausibility of the managers' "stories." This is true both in real estate and securities investment management.

The importance of the ex ante role of performance evaluation does not mean that the statistical problems with ex post inference can be ignored, however. The role of randomness in ex post returns, and also of ex post return measurement difficulties, especially in the private real estate asset class, should be carefully considered.19

This is particularly important in constructing investment management contracts that have incentive fee provisions based on performance evaluation (e.g., the use of performance "bogeys"). Too much randomness and lack of precision in the relationship between the ex post performance measure and the actual investment skill and diligence applied ex ante by the manager will act to demoralize or discourage good managers. The results could be perverse from the perspective of aligning the interests of the investor and manager.

For example, if too large a portion of the manager's overall fee is based on ex post performance relative to a bogey, or the bogey is defined inappropriately, then the best managers may shun investors who demand such contracts, or both sides may end up just "gaming" the incentive.20 This would not serve the interests of the manager, the client, or the industry as a whole in the long run.

The importance of minimizing randomness in the ex post return differential between the manager and an incentive contract bogey highlights the importance of selecting an appropriate benchmark for the manager's performance evaluation. The previous point about matching responsibility with authority is obviously important in this regard. It may also be important to base manager rewards on a less rigidly quantitative mechanism, so as to allow the combination of manager "stories" and ex post performance differentials relative to one or more benchmarks to be considered and weighed with a large dollop of common sense and wisdom.

CONCLUSION: NEITHER TO PRAISE NOR TO BURY?

This article has tried to step back and take a broad, deep, common-sense look at the question of investment performance benchmarking in the private real estate investment industry. Although, I admit this attempt may seem like a little too much of an exercise in apologetics for the NCREIF index, a warped paraphrase of Shakespeare's Mark Antony better describes my purpose. I have not come to praise the NCREIF index, but neither have I sought to bury it. The NCREIF index is not perfect, and it is not even as good as it can be, or should be, for purposes of serving the private real estate industry's need for a benchmark. But in this role, it is not terrible either, even as it is today, and NCREIF is striving to improve it.21

It was not so suddenly, indeed, it was only after a long struggle, not one but almost two decades ago, that NCREIF was founded and began to publish the NCREIF index. In that founding process, the U.S. real estate investment industry did something that was unprecedented and that remains very hard to do for an industry that is composed of highly competitive firms that thrive on private information. By giving up some of that private information to the NCREIF database, NCREIF's data-contributing member firms perform a public service for the common good of the private real estate investment industry as a whole, and indeed for the overall efficiency of capital allocation in the U.S. economy.22

Managers will often find it frustrating to work with the NCREIF index as a benchmark, and this is not only because no index can ever be perfect. Even in a perfect peer universe index, half the members will be beaten by the index in any given span of **historical** time, by definition and by construction. It is only natural to complain about that against

which we must compete.

The recent complaints that have been voiced against the NCREIF index as a benchmark have been largely constructive, and have stimulated NCREIF to make improvements. But, as I have attempted to explain or argue here, some of these complaints are not wellfounded, after all, after careful consideration. Some attacks against the NPI in its role as a benchmark amount to attacks against a straw man: nothing more or less than attacks against private real estate for being private.

No one can deny that private real estate is different from public securities. But that is not the fault of the NPI, nor of the real estate asset class. Nor is such difference a fault at all, in general. It is simply reality, and it will remain so, no matter how much the NPI is criticized, no matter how much the NPI is improved, and no matter what other brands or types of indexes are developed to represent private real estate in the U.S. Because of this reality, the private real estate asset class will never be able to produce the will-o-the-wisp of an index that is exactly comparable to the NAREIT or S&P 500 indexes in the public securities investment industry. But this does not imply that useful benchmarking cannot be done, or that useful information cannot be obtained, based on the NCREIF index.

Constructive criticism is welcome and needed in the real estate investment industry. But too much loud, unfounded complaining may needlessly undercut the credibility of the NCREIF index to the broader investor community, which largely lacks specialized familiarity with private real estate. Inaccurate or misleading criticism of the NCREIF index, or of any conscientiously and diligently constructed index in the private real estate asset class, could damage the viability of the entire asset class as a core component of the mixed-asset institutional **portfolio** by unjustifiably damaging real estate's credibility in comparison to other asset classes. Let's keep this in mind, and let's keep up the good work, together.

Is It a "Sample" or Is It a "Population?" The Statistical Nature of Benchmark Indexes...

Benchmark property-level real estate investment performance indexes such as the NCREIF Index are often spoken of in common parlance as being a "sample" of a larger underlying population of "all" the commercial properties of a certain "class" or type. Indeed, much basic research conducted using such indexes effectively treats property indexes this way, as they are taken, for example, to represent the performance of an entire "asset class" (which is generally somewhat vaguely defined). However, it is important to realize that, in the context of performance evaluation of investment managers, benchmark indexes such as the NCREIF index are not playing the role of statistical "samples" of some larger underlying population of properties. Rather, in this context they are meant to define an entire \population," namely, the population of all properties held by the relevant peer group of managers.

In the field of statistical inference, there is a large conceptual difference between a "population" and a "sample" of a population. One implication of this difference is that, if the index is viewed as a sample of a larger population, then, at least ideally, it should not be a **value**-weighted index (that is, an index constructed as the cross-sectional average of the individual property returns each period weighted by the relative magnitude of the property **value** within the overall population.) Instead, the ideal index from a statistical sampling perspective would be equally weighted across properties within each period. However, the NCREIF index is, quite intentionally, a **value**-weighted index. This suggests that the designers of the NCREIF index intended that it be viewed as a population rather than as a sample.

As noted, this conceptualization of the NPI as a population makes sense for benchmarking purposes, as peer universes are essentially populations. Nevertheless, NCREIF's new web site gives its members the

ability to easily construct equally weighted versions of the aggregate NCREIF index or any subindex within the NPI. This is a bow to the strong demand for use of the NPI as a sample of the U.S. commercial real estate asset class. However, it is important to distinguish this use of the NPI as a statistical research tool from its use as a manager performance evaluation benchmark. The ideal characteristics of an index are not the same for these two different uses. (See subsequent box.)

A Relevant Digression: The Industry Needs Better Research Tools...

This article has been about benchmarking private real estate investment manager performance, and the role of the NPI in this process. But of course, the private real estate investment industry has broader needs for investment performance indexes than just the benchmarking role. As the earlier box in this article notes, the NPI has always been regarded as somehow "representative" of the commercial property asset class in the U.S. In this role as a broader research tool; the NPI has served reasonably well throughout most of its history as what might be called a \proxy" for the investment performance of the private commercial real estate asset class. (The term "proxy" refers to a Less formal and less rigorously defined representative of a population than the term "statistical sample.") Nevertheless, it is also important to recognize that the needs of the broader research community for investment performance indicators are more varied, and at times more stringent and demanding, than the needs of the investment industry for a manager performance benchmark.a

Rigorously speaking, it is probably correct to note that the design and construction of the NPI falls much farther below "optimality" for a role as a broad research tool representing the commercial property asset class than it does for a role as a private real estate investment manager performance benchmark. From an academic perspective, an ideal research index representing the commercial property asset class would be defined and constructed as a scientifically stratified "statistical sample," not as a \population" of all (and only) the properties that happen to be owned by NCREIF members. The NPI's problem as a sample of the commercial real estate asset class is not so much that it is small in size, but that it is not a scientifically constructed **random**gt; or representative sample.

Of course, in order to design an ideal stratified sample of properties, one would first have to explicitly define the population of properties one is trying to represent, at least conceptually (e.g., is it all properties in the country over $10 million in **value**gt;? Is it only income-producing properties in official MSAs? etc.). Once the underlying population of properties is defined, it would not be necessary to include the entire population within the sample. One of the great principles of statistical inference is that samples derive their power not from the percentage of the underlying population they include, but only from their absolute size. Furthermore, it does not require a very large absolute sample size to obtain considerable power, as statistical inference power is generally a function of the square root of the sample size. Once the population and its key subcategories are defined, a stratified sample including a statistically useful minimum number of properties in each category would ideally be identified as **random**gt; or representative example properties in each category.

NCREiF is currently exploring something like this in its so-called "NCREIF 500" or "NCREIF 1000" project. But, of course, NCREIF is limited to the sample of properties owned by its data-contributing member firms. Ideally the real estate investment research community would like something broader. And we face an historic opportunity at the present time to develop such a "super-- index" of property-level performance information. The development of the Internet and "e-commerce" in the commercial property market will make information on property prices, rents, leases, and other topics much more available and centralized, and in a readily accessible electronic format.\ The development of repeated measures regression (RMR)-based statistical methodologies also greatly facilitates the production of useful indexes from transaction price information.d The

increasing centralization and public accessibility of information may enable individual for-profit firms to produce performance indexes that are as broad, or broader, than the NCREIF Index. Based on its mission as a non-profit research and information dissemination association for the real estate investment industry, NCREIF should welcome and support, as well as contribute to, such developments.

The real estate investment research community includes government and academia, as well as the private sector. It encompasses branches of the industry that deal with public as well as private assets, debt as well as equity. In addition to NCREIF, relevant industry associations include PREA, NAR-EIM, NAREIT, MBAR, CIREI, and ACLI, among others. Interested government agencies include the FRB, SEC, BEA, and the Census, among others. And academic or research organizations would include AREUEA, ARES, and RERI. The time is ripe to engage a broad segment of this community to look at commercial property investment performance index production from a far-ranging "clean-slate" perspective to scope out a new paradigm for the production of the commercial property research tools appropriate for the new century.

ENDNOTES

a See Fisher and Geltner [2000] for additional discussion of the NPI and its use and development for various research roles.

b For some research purposes, there are other problems as well, such as the nature and amount of property performance and **valuation**gt; information gathered, including the nature of the appraisal information. (For example, see Fisher and Geltner (2000].)

c XML. technology can greatly facilitate the necessary information-sharing. See the "DataConsornum" initiative at www.dataconsortium.org. The DataConsortium is an independent, open-forum initiative that grew out of activity in the NCREIF Research Committee.

d See Fisher and Geltner (2000] for firther description of the RMR methodology.

ENDNOTES

'Normally, the time-weighted average return is used (in particular, the geometric average), because at the **portfolio**gt; or fund level (as distinct from the property level), the investment manager does not usually have much control over the timing of capital flow. Time-weighted average returns are neutral with respect to capital flow timing, while dollar-weighted returns reflect the effect of such timing. If the manager substantially control the timing of capital flow, then the IRR might be a more appropriate measure of the manager's performance. However, benchmark performance should always be measured by a time-weighted return. The benchmark is meant to be neutral with regard to performance actions. In the case of capital flow timing, such neutrality is represented by the timeweighted average return. The geometric mean is typically used rather than the arithmetic mean. This is because performance evaluation is a backward look at history rather than a forecast

ing or sampling exercise in which the statistical properties of the arithmetic mean would be useful, and because the geometric mean is not sensitive to the volatility in the periodic returns. (The arithmetic mean is greater the more volatile the returns.) Also note that at the **portfolio** or fund level being addressed here (as distinct from the property level), the periodic HPRs are normally computed as **value**-weighted (cross-sectional average) returns within each period. That is, within each period the return to each property is not weighted equally but by the relative **value** of the property

within the **portfolio**. This type of **value**-weighting is necessary to enable the HPR each period to reflect the actual return to the **portfolio** within that period. The focus here is on **portfolio**-level returns, rather than viewing each property as a sample of the manager's performance.

zThe placement of Manager A's and Manager B's performance evaluations together in the Exhibit is done merely for illustrative purposes. It is not meant to imply that there should be a "head-to-head" competition between these two managers. Such direct comparisons would normally only take place between managers within the same asset class or segment (or "style") of investments, to avoid an "apples versus oranges" comparison.

'The term "active management" in the investment management industry has traditionally been defined from a public securities market perspective. In that context, of course, active management does not imply operational management of the investment assets, as this is not possible for investment managers in the securities industry. But in a broader sense, active management refers to the attempt to "beat the market," or to outperform the "average" or "typical" **portfolio** of assets, through the use of whatever means are at the disposal of the manager. In the case of securities investment managers, this is limited to the use of investment research and the employment of active trading strategies or target allocations. These types of activities, broadly defined, may also be employed by private real estate investment managers, although in a more sluggish asset market, dealing with assets that are primarily income-generators rather than growth plays, and in an environment of high transaction costs. But in addition to these traditional active management tools, private real estate investors can, indeed must, use their control over property-level operational management of the assets to try to improve their investment performance.

4While a peer universe contains, in principle, all the managers of a certain type that are currently active, there is a potential "survivorship bias" problem, which is often raised as a criticism of peer universes as compared to passive indexes for performance benchmarks in the securities industry. Survivorship bias occurs when funds or managers who have failed or been taken off the market disappear from the peer universe. However, there is some reason to believe that, at least through the 1990s, survivorship bias has not been a serious problem in the NCREIF index. For one thing, the NCREIF index is a property-level, as opposed to a fund-level index. Also, real estate

managers and funds tend to be more stable than stock market funds. Indeed, the number of investment managers and properties covered by the NCREIF index remained fairly stable even through the major real estate downturn of the early 1990s. One reason is that real estate funds and managers often do not fail outright, but are acquired by other funds or managers, which typically results in the properties remaining in the index or exiting on the basis of transaction prices or updated appraised values that largely reflect recent **value** losses.

5Note, however, that property-level performance is "gross" of management fees and other "fund-level" or "manager-level" impacts on the net returns actually received by the investor client. This is also true for passive index benchmarks in the securities industry, such as the S&P 500. Of course, what really matters to the investor are the net returns received at the investor level.

6In an extreme case it may be that the institutional investor peer universe owns virtually all the property in a given market segment (probably a rather narrowly defined market segment). In that case, the peer universe becomes virtually identical to the market, which may be what people have in mind as a "passive index" in the private real estate asset class. However, even in this case it will not be possible for the manager or the investor to exactly replicate or invest in the index as an alternative to working with an investment manager, for the reasons noted previously. In short, the notion of a "passive index" does not have the

same meaning or rationale as a benchmark in the private real estate asset class as it does in the public securities investment industry.

'In fact, because of the lumpy nature of investment in the private property market (because whole assets are traded, not shares), even constant segment allocation weights in the manager's benchmark might force the manager into placing at least a slight "bet" against his benchmark, whether he wanted to or not. Even if the manager was able to perfectly match the benchmark's initial segment allocation weights, over time the weights in the manager's **portfolio** would change merely by virtue of differential growth in property values within the manager's **portfolio**, even if the manager were completely passive in the sense of holding a constant set of properties.

BFor example, if the manager and client have agreed that the manager's benchmark should be 50% office and 50/ industrial property, then the NPI office and industrial subindexes can be used, each with a 50% weighting, to construct the manager's benchmark.

9For example, neither ex post nor ex ante returns within the private core commercial property asset class are positively correlated with property market segment "betas," whether such betas are measured with respect to the NCREIF index, or the NAREIT index, or the stock market, or any other known risk benchmark, and this lack of an empirical CAPM-- based risk/return relationship within the private real estate

core holds even after correcting for "smoothing" and "lagging" within the property returns.

"In other words, there is perhaps less theoretical reason, within the private real estate asset class as compared to the stock market, to view the "market **portfolio**" as reflecting an "optimal" segment weighting in a normative sense. Furthermore, from a practical perspective, it is difficult to precisely define or estimate market **portfolio** weights within the commercial property asset class. For an example of an attempt to quantify the geographic allocation across regions in the U.S., see Mahoney, Malpezzi, and Shilling (2000.

"Keep in mind, however, that specialized local expertise is necessary to enter into specific real estate market segments. One should be wary of defining a benchmark that encourages a manager to go into territories or property types where he lacks familiarity. If market weightings are to be used, they should be applied only across the segments in which the manager has sufficient expertise and resources to operate effectively.

'12Another way of putting this is as follows. Suppose a manager beat his benchmark by 50 basis points over the past three years. What does that imply about how we can rationally expect this same manager to perform relative to the same benchmark over the next three years?

'3The longer the observation sample, the lower the volatility, and the higher the correlation, the easier it is to accurately infer the true difference from the sample drawings. During the 1980s and 1990s, the quarterly volatility of the S&P 500 was about 8%, that of long-term Government bonds was about 6%, while the volatility of the NCREIF Property Index (NPI) was only about 2%. However, the NPI exhibits very strong positive autocorrelation in its quarterly returns. As a result, in a time sample of, say, twelve quarters, the NPI does not really provide twelve independent observations of the return, in contrast to an index like the S&P 500, for example. If we delagged the NPI so that its quarterly periodic returns were like independent observations (a zero-autocorrelation version of the NPI), then the volatility of the NPI would easily exceed 5/ per quarter. Furthermore, 90% is much greater correlation than one can usually obtain between a manager's **portfolio** and an appropriate benchmark. For example, in Myer and Webb [1993], the quarterly HPRs of professionally managed private real estate portfolios were regressed onto the NPI over twenty-nine quarters. The average RZ of these regressions across the forty-seven portfolios they examined was 0.25. This implies an average correlation of SQRT(0.25), or 50%.

'''Note that for this type of inference we use the arithmetic average

return rather than the geometric average. The computations behind this conclusion are as follows. If 6_ is the manager's volatility, 6B is the benchmark's volatility, and C is the correlation between the two, then the following formula gives the standard deviation of the difference between the manager's return and the benchmark's return, r., -rB,,:

sigma
sub i -B
= (sigma
sub i
sup 2
+ sigma
sub B
sup 2
- 2Csigma
sub i
sigma
sub B
)
sup (1/2)
= [0.05
sup 2
+ 0.05
sup 2
-2(0.9)(0.05) (0.05)]
sup (1/2)
=2.24%

Dividing this by the square root of N - 1, where N is the number of observations (twelve in this case), gives the standard error of the observed average difference:

Std.Err. = sigma
sub i-B
/((N - 1)
sup (1/2)
) = 2.241/o/3.32 = 0.675%

Two standard errors are required for "statistical significance" (95% confidence), which would be 2 X 0.675%, or 1.35% per quarter. This equates to about 5.5% annualized. Note also that, although the arithmetic average return is, in principle, more appropriate for this type of statistical comparison, it is still necessary to consider the effect of volatility on the arithmetic average return if the two performance series being compared display different volatility (which is not the case in the present numerical example). One (somewhat crude or informal) way to adjust for the effect of volatility on arithmetic mean returns is to simply use the geometric mean returns of the two series being compared, even though the confidence limits are computed as above based on arithmetic means.

'sln the forty-seven portfolios they examined, Myer and Webb [1993] found only sixteen whose returns exceeded the NPI with statistical significance. Furthermore, as noted, such indications of statistical significance assume that the underlying relationship between the manager and the benchmark is constant over time, and that the realized holding-period returns to both the manager and the benchmark can be observed without error each period. In reality, these assumptions are not true even in public securities markets, but especially not in the private real estate market. Return measurement errors muddy the waters, and may require even larger spreads before one can conclude with confidence that a manager actually beat his benchmark by a sufficient margin to rule out luck as the cause.

'6The "warning label" presented here is taken from the First Edition [1998] of the Australian Investment Performance Measurement & Presentation Standards, published by the Property Council of Australia.

"The biggest incentive for the manager may simply be the prospect of having the management contract renewed for another three (or more) years.

'18Incidentally, this is as true for a peer universe benchmark as it is for a passive index, so from this ex ante incentive perspective, peer universe-based benchmarks are generally as functional as passive indexes. Indeed, for this ex ante purpose it suffices that the benchmark merely be such that the manager substantially improves his ex ante likelihood of beating the benchmark if he works harder (smarter) for the client during the period of the management contract. Note that this does not require a "statistically perfect" benchmark.

'9Temporal lag bias in real estate returns may not be a prohibitive problem in performance evaluation, as such bias may

largely cancel out in the performance differential between the manager and his benchmark. It is important, however, to ascertain whether the degree of lagging is similar in the manager's reported returns and those of the benchmark. If the lag is similar, then smoothing or lagging will probably not affect the performance evaluation much. However, **random** noise (purely **random** return measurement error as distinct from temporal lag bias) may still pose a serious measurement problem in the performance differential. Such measurement error is an additional source of randomness beyond the underlying volatility in the true returns across time. Noise will be a particular problem if either the manager's **portfolio** or the benchmark contains a relatively small number of individual properties or appraisers.

2"In the extreme, if there is too much **random** measurement error in quantifying the performance differential between the manager and the benchmark, then poor managers will have almost as good a chance of earning an incentive fee as superior managers.

2"For example, consistent with its mission, NCREIF is in the process of adding fund-level data that will include, for the first time, the major opportunity funds that manage most institutional investment in "non-core" property types, such as development projects, land, and hotels. This will improve the ability of the NCREIF index to be representative of broader segments of the investment manager universe. NCREIF is also actively researching the development of several new types of index products, such as an index of properties held only in open-end funds, and an index consisting of a relatively static sample of "typical" properties that would be explicitly identified, and for which more timely appraisals and more complete operating information would be collected and reported. (This latter might be called the "NCREIF 1000," as it might consist of approximately 1,000 out of NCREIF's approximately 2,500 properties.) NCREIF has also launched an interactive web site for its members that allows easy construction of custom benchmarks from the NCREIF database.

22Of course, this is not a service that is without private rewards for NCREIF members, both direct and indirect, such as access to the above-noted interactive web site.

REFERENCES

Fisher, J., and D. Geltner. "De-Lagging the NCREIF Index: Transaction Prices and Reverse-Engineering." Real Estate **Finance**, 17, 1 (Spring 2000), pp. 7-22.

Mahoney, J., S. Malpezzi, and J.D. Shilling. "Implications of Income Property Stock Data for Real Estate Investment **Portfolio** Location." Real Estate **Finance**, 16, 4 (Winter 2000), pp. 53-66.

Myer, F., and J. Webb. "Return Properties of Equity REITs, Common Stocks, and Commercial Real Estate: A Comparison." Journal of Real Estate Research, 8, 1 (Winter 1993), pp. 87-106.

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION Required by 39

U.S.C. 3685

1. Publication Title: Real Estate **Finance**

2. Publication No.: 726-370

3. Filing Date: February 15, 2000

4. Issue Frequency: Quarterly

5. No. of Issues Published Annually: 4

6. Annual Subscription Price: $240

7. Complete Mailing Address of Known Office of Publication (Street, City, County, State, and ZIP+4) (Not Printer): Institutional Investor, Inc., 488 Madison Avenue, New York> NY 10022

8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not Printer): Same as above

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher. Gauri Goyal, Institutional Investor, Inc., 488 Madison Avenue, New York, NY 10022. Editor: Mike Miles, Fidelity Management & Research Co., 82 Devonshire Street, F7E, Boston, MA 02109-3614. Managing Editor: Profs. Norman Miller/David Geltner, Dept. of **Finance**, University of Cincinnati, 414 Carl H. Lindner Hall, Cincinnati, OH 45221-0195.

10.Owner (If owned by a corporation, its name and address must be stated and also immediately thereafter the names and addresses of stockholders owning or holding 1% or more of the total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual, must be given. If the publication is published by a nonprofit organization, its name and address must be stated.):

Associated Newspapers Holdings Ltd.
Northcliffe House
2 Derry Street
London W8 STT
Prudential Client (MSS) Nominees Ltd.
Manner House
Pepys Street
London EC3N 4DA
Lloyds Bank SF Nominees Ltd.
Lloyds Bank Securities Services
Hays Land House
1 Hays Lane
London SE1 2HA
State Street Nominees Ltd.
One Canada Square
London E14 5AF
CO Nominees Ltd.
Lloyds Bank Securities Services
Hays Lane House
1 Hays Lane
London SE1 2HA

11.Known bondholders, mortgagees, and other securityholders owning or holding 1% or more of total amount of bonds, mortgages,
or other securities (if there are none, so state): None

12. For completion by nonprofit organizations authorized to mail at special rates. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Not applicable

13. Publication Name: Real Estate **Finance**

14. Issue Date for Circulation Data Below Fall 1999

15. Extent and Nature of Circulation:

DAVID GELTNER is the Power professor of real estate at the University

of Cincinnati and serves as NCREIF's academic advisor.

Special Features: Illustration; Table
COMPANY NAMES: National Council of Real Estate Investment Fiduciaries
Classification: 9130 (CN=Experimental/Theoretical); 8360 (CN=Real estate); 3400 (CN=Investment analysis & personal finance); 9190 (CN=United States)
Descriptors: Benchmarks; Studies; Real estate appraisal; Price indexes; Performance evaluation; REITs
Geographic Names: United States; US
PRINT MEDIA ID: 13978

36/9/2 (Item 2 from file: 268)
00377341 46983967
Credit unions and the common bond
Emmons, William R; Schmid, Frank A
Review - Federal Reserve Bank of St. Louis , v 81 , n 5 , p 41-64 , Sep/Oct 1999
CODEN: FRBRDV **Document Type:** Periodical; Feature **ISSN:** 0014-9187 **Journal Code:** FSL
Language: English **Record Type:** Fulltext

Word Count: 10,537

Abstract:

A model of credit union formation and consolidation is developed and simulated in order to examine the common-effects of common-bond restrictions on the performance of credit unions.

Text:

Cooperative **financial** institutions have their roots in 19th century Europe, appearing first in the United States during the early 20th century Cooperative **financial** institutions are ubiquitous in both developed and developing countries today, posing something of a puzzle in the former group of countries where one might have expected corporate **financial** institutions with professional management and sophisticated capital-market oversight to have displaced them. This has not occurred, however, as some groups of cooperative **financial** institutions in developed countries are holding steady or even increasing their market shares. In the United States, the most prominent types of cooperative **financial** institutions today are mutual savings and loans, mutual savings banks, mutual insurance companies, and credit unions.

Credit unions are regulated and insured **financial** institutions dedicated to the saving, credit, and other basic **financial** needs of selected groups of consumers. By law, credit unions are cooperative enterprises controlled by their members-under the principle of "one-person one-vote." In addition, credit union members must be united by a "common bond of occupation or association, or (belong) to groups within a

well-defined neighborhood, community, or rural district" (Supreme Court, 1998, p. 2, quoting from the Federal Credit Union Act of 1934).

Despite the rather low profile and mundane operations of the vast majority of credit unions, these institutions have long been a source of controversy in the United States. Public awareness of this long-simmering debate was piqued recently by a Supreme Court case pitting commercial banks against credit unions and their federal regulator (Supreme Court, 1998). The Court found in favor of banks in this case, ruling that the federal credit-union regulator, the National Credit Union Administration, must cease granting federally chartered credit unions the right to combine multiple common bonds (fields of membership) within a single institution. Less than six months later, however, President Clinton signed into law new legislation that essentially reversed the Supreme Court's ruling.

This paper provides background on credit unions and the debate they have spurred in the United States. In addition, we present new evidence relevant to the credit-union debate concerning fields of membership (common bonds). Our **analysis** is based on a theoretical model of credit-union formation and consolidation. Using an extensive dataset and a nonlinear empirical approach, we find that creditunion participation rates generally decline as the group of potential members becomes larger, holding all else equal. That is, the larger the pool from which a singlegroup credit union can draw, the less effective it is in attracting members.

We also provide new evidence on two more general banking policy issues. First, we find evidence to support the structureconduct-performance paradigm of local banking competition. This is the prediction, derived from theoretical considerations, that more concentrated markets ultimately lead to higher prices and lower quantities. Policymakers have used this paradigm extensively when justifying intervention in the market for corporate control in **financial** services. Using the Herfindahl index calculated for local bank deposit market shares as a measure of local market structure, we find that higher levels of market concentration are associated with higher participation rates at credit unions. This is consistent with the notion that banking competition is weaker in more concentrated markets, which increases the attractiveness of credit unions.

The second banking policy issue we address is that of possible scale economies among **financial** institutions. Our empirical results indicate that credit unions generally encounter significant scale economies, whether scale is measured by the log of total assets or by the log of the number of credit-union members. The latter finding, however, applies only to relatively large credit unions.

It is important to point out several limitations of this study As in all empirical investigations, we can describe relationships in the existing data but we cannot predict exactly how these relationships would appear under a different set of operating conditions. For example, an extended period of growth by many credit unions could alter the extent of scale economies that exist. Similarly, significant changes in credit-union regulation might result in different empirical regularities than those identified here. It also is important to keep in mind that we abstract from managerial agency problems in credit unions in this article (see Emmons and Schmid, 1999, for an extensive discussion of this issue). Finally, it is hazardous to draw conclusions about public policy toward credit unions on the basis of this rather narrowly focused investigation. We hope to provide insights into the effects of common-bond requirements, not to provide a comprehensive framework for evaluating competition in the **financial**-services sector as a whole.

The paper is organized as follows: The first section provides some institutional and **historical** background on credit unions, while the second section outlines the current credit-union debate in the United States. The third section develops a theoretical model of credit-union formation and consolidation. The model stresses the countervailing influences on participation rates of (1) scale economies in production, and (2) decreasing withingroup membership affinity as a credit union grows. The

model provides intuition for why the number of common bonds within a credit union might be important for their formation and growth. The third section also describes a simulation of the theoretical model that can be used to generate some comparative-static results. The fourth section briefly describes the dataset and the econometric methods we employ in analyzing federally chartered occupational credit unions. The fifth section presents our empirical results, and the sixth section draws conclusions. An appendix describes the data we use.

This section provides some institutional background to help motivate the theoretical and empirical analyses later in the article. The key points this section seeks to illuminate are the restrictions on credit-union expansion and the arguments that have been made to support or oppose these restrictions. The sections that follow investigate the extent to which the common-bond requirement acts as a binding constraint on credit-union operations.

Credit unions numbered 11,392 at year-end 1996, serving some 70 million individual members (U.S. Treasury, 1997, p. 15). At the same time, there were 11,452 commercial banks and thrift institutions (savings and loan associations and mutual savings banks). Credit-union assets were only $327 billion, compared to $5,606 billion held by commercial banks and thrifts (U.S. Treasury, 1997, p. 21). A more direct standard of comparison might be community banks and thrifts, however. At year-end 1996, there were 7,049 community banks and thrifts (defined as all federally insured banks and thrifts with less than $100 million in assets) holding combined assets of $324 billion (U.S. Treasury, 1997, p. 21). A comparison of credit unions and community banks and thrifts is particularly meaningful because institutions of both types are relatively focused institutions, and hence, are unable to grow beyond certain limits. For example, a single-employer occupational credit union is authorized to serve only the employees of the sponsoring firm and their immediate relatives, who may total no more than a few hundred people. A community bank or thrift may operate in only one geographical area. In addition, credit unions are restricted in the types of **financial** services they may provide, with traditional consumer **financial** services at the core of virtually all credit unions' activities. Community banks and thrifts may offer a similar array of services.

Both federal and state agencies grant credit-union charters. Regardless of the type of charter they hold, the deposits (or technically, "shares") of virtually all credit unions are now federally insured by the National Credit Union Administration (NCUA). Federal credit unions are regulated by the NCUA while state-chartered credit unions are regulated by an agency of the chartering state.

Of the 7,068 federally chartered institutions at year-end 1996, about three quarters were occupational credit unions (U.S. Treasury, 1997, p. 19).1 In an occupational credit union, one or more firms sponsor a credit union, sometimes providing office space, paid time off for volunteer workers, and perhaps other forms of support. The remaining federal credit unions were either single-group associational or community credit unions, or multiple-group credit unions with predominantly associational, community, or more than one type of membership (i.e., several groups that span the usual classifications).

By size, most credit unions (65 percent of federally insured institutions) had less than $10 million in assets (U.S. Treasury 1997, p. 19). Large credit unions exist, however, and they are an important part of the sector. For example, the 11 percent of credit unions with more than $50 million in assets (1,284 institutions) accounted for 74 percent of total credit-union assets.

Credit unions play a limited role in the U.S. **financial** system, catering to the basic saving, credit, and other **financial** needs of well-defined consumer groups. More than 95 percent of all federal credit unions offer automobile and unsecured personal loans, while a similar proportion of large credit unions (more

than $50 million in assets) also offer mortgages; credit cards; loans to purchase planes, boats, or recreational vehicles; ATM access; certificates of deposit; and personal checking accounts (U.S. Treasury, 1997, p. 23). Very small credit unions typically offer a limited range of services, are staffed by membervolunteers, and are likely to receive free or subsidized office space. Larger credit unions offer a broader array of services. They may employ some full-time workers, including the manager, and are more likely to pay a market-based rent for office space.

Historically, members of credit unions were drawn from groups that were underserved by traditional private **financial** institutions; these consumers tended to have below-average incomes or were otherwise not sought out by banks. While credit-union members today still must share a common bond to be eligible for membership, the demographic characteristics of credit-union members have become more like the median American. While only I percent of the U.S. adult population aged 18 or over belonged to a credit union in 1935, some 33 percent of the adult population had joined by 1989 (American Bankers Association, 1989, p. 29). Subsequent strong growth of new credit-union charters has increased that proportion.2

According to a credit-union survey in 1987, 79 percent of all Americans who were eligible to join a credit union had done so (American Bankers Association, 1989, p. 29). Given the prominent role of occupational credit unions, a majority of members are in the prime working ages of 25-44 (American Bankers Association, 1989, p. 30). Perhaps surprisingly, given the origins of credit unions, current members are overrepresented in upper-middle income strata, defined as household incomes between $30,000 and $80,000 in 1987. Overall, it appears that credit unions, banks, and thrifts are more direct competitors today than when credit unions first appeared.

The predecessors of American credit unions were cooperative banking institutions of various sorts in Canada and Europe during the 19th century. The first credit union in the United States was formed in Manchester, New Hampshire, in 1909 (U.S. Treasury, 1997, p. 15). Soon thereafter, Massachusetts created a charter for credit unions. The credit-union movement swept across the United States from there, meeting with particular success in the New England and upper Midwestern states.

These early cooperative **financial** institutions often had a social, political, or religious character in addition to their explicit economic function. While the social and political aspects of the cooperative movement were acknowledged and accepted by the United States Congress, the Federal Credit Union Act (FCUA) of 1934 was focused more narrowly on the economic potential of credit unions.

The legislation itself was modeled closely on state credit-union statutes that had appeared during the early decades of the 20th century in the Northeast and upper Midwestern states. The FCUA clearly reflected Congressional intent to create a class of federally chartered **financial** institutions that would operate in a safe and sound manner: ... the ability of credit unions to "come through the depression without failures, when banks have failed so notably, is a tribute to the worth of cooperative credit and indicates clearly the great potential **value** of rapid national credit union extension." (Supreme Court, 1998, p. 17, citing the FCUA, S.Rep. No. 555.)

The likelihood that federal credit unions would serve consumers not served by banks was an additional element in Congressional deliberations:

Credit unions were believed to enable the general public, which had been largely ignored by banks, to obtain credit at reasonable rates. (Supreme Court, 1998, p. 17.)

Partly because credit unions are mutual associations, they were not subjected to federal taxation as were shareholder-owned commercial banks and thrift institutions. Mutuality cannot be the only reason why credit unions are not taxed, however. Other mutually owned enterprises are subject to taxation. As for the benefits of tax exemption, credit unions (or any other firm) could avoid paying taxes by paying out all "profits" to members

in the form of lower borrowing rates or higher deposit rates. The real importance of the tax exemption is that credit unions can retain earnings tax free. Advocates argue that this is justified because credit unions cannot raise equity in a public offering, so they must be able to build capital internally.

It is clear from the legislative history surrounding the passage of the FCUA in 1934 that Congress saw the common-bond requirement as critical to the success of credit unions:

The common bond requirement "was seen as the cement that united credit union members in a cooperative venture, and was, therefore, thought important to credit unions' continued success. ..... "Congress assumed implicitly that a common bond amongst members would ensure both that those making lending decisions would know more about applicants and that borrowers would be more reluctant to default." (Supreme Court, 1998, pp. 17-18, citing 988 E2d, at 1276.)

The subsequent history of credit unions in the United States largely has fulfilled the promise envisioned by Congress in 1934. Credit unions have grown and spread across the country Although hundreds of individual credit unions failed during the 1980s and early 1990s, the National Credit Union Insurance Fund (NCUSIF, formed in 1970) avoided accounting insolvency-in marked contrast to the Federal Savings and Loan Insurance Corporation and the Bank Insurance Fund of the Federal Deposit Insurance Corporation (Kane and Hendershott, 1996). Credit unions control a small but growing share of household deposits, and some of our empirical results indicate that they may play a role in maintaining a high level of retail banking competition in some local markets.

The special status and comparative success of credit unions in recent decades, coinciding as it has with a period of stress on thrift and commercial-banking institutions, has led to political conflicts between advocates of credit unions and banks.

This conflict reached its high point in a series of court decisions culminating at the U.S. Supreme Court in October 1997. The particular case at issue involvect the AT&T Family Credit Union and the NCUA5 interpretation of the 1934 FCUA allowing multiple common bonds of membership. Brought by several banks and the American Bankers Association, the case was ultimately decided in February 1998 (on a 5-4 decision) in favor of the banks who sued to stop the NCUA from granting more multiple-group credit-union charters. The bankers' victory was short-lived, however, as Congress almost immediately drafted new legislation that enables credit unions to continue growing much as before-including multiple common bonds within a single credit union.

The shaded insert summarizes the key provisions of the Act.

Attacks on credit unions have come from a wide range of viewpoints, the proponents of which have wielded sometimes contradictory arguments. Some of the arguments used in the recent Supreme Court decision concerning the role of the common-bond requirement in credit unions reflect the unsettled nature of the debate. We focus on two strands of the credit-union debate here, namely the arguments stressing inefficient governance structures on the one hand and unfair competition on the other.

Some have argued that credit unions are inherently inefficient clue to their onemember one-vote governance structure. One might expect decision-making in a credit union to be of poor quality due to a lack of professionalism (i.e., volunteer managers and workers), free-riding of members in monitoring the management, and weak incentives for members to intervene when action is needed to correct specific problems or deficiencies.3 According to this argument, credit unions may waste scarce resources and they may eventually impose significant costs on individual sponsoring firms or the economy as a whole.

The second prominent line of argument aimed at credit unions takes a nearly opposite view of their organizational effectiveness. This view presumes that credit unions operate efficiently enough to offer consistently better terms on savings and credit services than those offered

by commercial banks and thrifts. Bank and thrift managers and owners often present this point of view in public discourse.

To be sure, those arguing that credit unions represent unfair competition ascribe some or all of their competitive advantages to subsidies such as their taxexempt status or sponsor subsidies rather than inherent efficiency.

Proponents of the first view-that credit unions are inherently inefficienthave a difficult time explaining why the number of credit unions and credit-union members continues to grow, and why members express high levels of satisfaction with the services they receive. if most credit unions were very inefficient, one might expect their members to become disaffected and their role in the **financial** system to diminish over time.

On the other hand, proponents of the second view-that credit unions are unfair competitors due in part to subsidiescannot explain easily why credit-union sponsors and governments are such strong supporters of credit unions. It is hard to understand how large net subsidies could be delivered to credit-union members over time without more opposition arising from constituencies that might be paying the subsidies, such as shareholders or employees who do not belong to their firm's occupational credit union, or taxpayers who belong to no credit union. In fact, the most vocal complaints about alleged subsidies for credit unions are heard from banks and thrifts, whose resentment of credit-union competition could be expected even if there were no subsidies flowing to credit unions.

Ironically, the juxtaposition of these two lines of attack against credit unions appeared in the argumentation of the Supreme Court majority that decided the AT&T Family Credit Union case in favor of commercial banks. At one point in its opinion, the majority cited the legislative history surrounding the 1934 Federal Credit Union Act as support for the view that credit unions are a fragile-even flawed-type of institution, reasoning that: Because, by its very nature, a cooperative institution must serve a limited market, the legislative history of Section 109 demonstrates that one of the interests "arguably... to be protected" by Section 109 is an interest in limiting the markets that federal credit unions can serve. (Supreme Court, 1998, footnote 6, pp. 8-9.) Thus, a credit union would become inefficient if it grew beyond its "limited market," as defined by its common bond.

At a different point in its opinion, however, the majority accepted the argument that credit unions with multiple groups of members would be more formidable competitors to banks and thrifts than single-group institutions. The majority argued that an expansive interpretation of the 1934 Act "would allow the chartering of a conglomerate credit union whose members included the employees of every company in the United States (1998, p. 4)." In other words, credit unions would overwhelm banks and thrifts unless otherwise constrained.

The irony is, of course, that the argumentation based on the reductio ad absurdum of a hypothetical "conglomerate credit union" did not mention the legislative history of the 1934 Act, which had essentially predicted that such a huge credit union would not have been a safe and sound **financial** institution, nor consequently a viable one in the long run.

How should policymakers think about credit unions? Are they relics of a bygone era, propped up by subsidies and distorting **financial** -sector competition? Or, are they efficient and focused **financial** institutions that could, if unleashed, eventually dominate some or all of the retail **financial** landscape? We do not seek to answer these emotionally charged questions directly. Instead, we focus on the more limited question of what effect the common-bond restriction exerts on creditunion formation and consolidation. In a sense, we are merely attempting to answer the question, "Does the common-bond requirement constrain the existence or growth of credit unions?" We hope that our insights may contribute to a better understanding of the larger

policy questions mentioned above.

In this section we present a model of credit-union formation and consolidation. We then describe the results of a simulation of the model. Subsequent sections of the paper discuss testable hypotheses emerging from the model, the data we examine, and empirical results.

We take for granted that credit unions typically are small; that they encounter operating economies of scale as they expand from a very small base of members and assets; and that they face direct competition from banks. The key trade-off we model is between decreasing affinity among members as the potential membership grows (i.e., as a given common bond is extended to more people)-making a credit union less effective-versus the increasing scale economies that come with a larger base of members and assetsmaking a credit union more effective. We show that the ability of credit unions to expand by adding multiple common bonds to their membership affects this trade-off in an important way.

We examine a Hotelling (1929) economy consisting of a "city" that lies on a straight line of unit length. The city's length is covered by a continuum of households. The location of each household corresponds to its preferences for banking services. In particular, each household demands exactly one unit of banking services but the nature of desired services differs among households. Preferences in the real world are, of course, multidimensional, encompassing tastes for different menus of **financial** services, different levels of service, or different locational preferences. We assume for the sake of simplicity, however, that a household's preferences for banking services can be represented in terms of a single index running from zero to one. Figure I depicts the linear-city model.

Because we are interested only in the formation and consolidation of credit uni6ns, we assume that credit unions are scarce (or differentiated) while commercial banks are ubiquitous (or uniform). In other words, consumption of credit-unionprovided **financial** services takes place at the point on the unit **interval** where a credit union is located, while commercial-bank services are available at a fixed price at any point on the line. This assumption makes household preferences critical for the existence of and participation in credit unions while maintaining the realistic assumption that commercial banks provide an alternative to credit unions (and vice versa). .

We assume that the entire city (i.e., ever), point on the line) is covered by at least one household and at most two households. Without loss of generality, we assume that all points covered by two households are arrayed continuously from zero upward towards, but potentially short of, one on the unit **interval**. For expositional purposes, we will refer to the households that inhabit the completely covered zero-to-one **interval** as being above the line and all others as below the line. Thus, two households that possess identical locations (preferences) are said to be "back-to-back" households.

Households are further grouped by affinity, or common bonds. For tractability, we discuss occupational common bonds and limit the number of employers in the economy to three. Each household located above the line contains an employee of either firm A or firm B (but not both). Because all households in employee group A share a common bond, they are located in a contiguous segment of the line that does not overlap the domain of employee group B. All households below the line contain employees of firm C. Each employer may sponsor a credit union, although, as we will see, not all will do so. We examine two periods (or regimes), differentiated according to the permissibility of forming credit unions with multiple common bonds. All households are born at the start of period I and live through the end of period 2. Each household needs to consume one unit of banking services in each period. These services can be provided by an occupational credit union or by a bank in either period.

At the beginning of the first period, households find themselves arrayed along the city's unit **interval**. The lengths of the

firm-A and firm-C segments are distributed as uniform **random** variables on the 10, 11 **interval**. The length of the firm-B segment is one minus the length of the firm-A segment.

We simulate the model by drawing repeatedly (10,000 times) a set of five uniformly distributed **random** numbers from the [0, 1] **interval**. The first draw determines the length of the segment containing households with an employee of firm A. Recall that the length of the segment containing firm-B households equals one minus the length of the firm-A segment. The second draw determines the length of the segment containing households with an employee of firm C. This determines the length of the line segment that is covered by two households. The last three **random** numbers enter the three (potential) credit unions' cost functions as stochastic elements of their fixed costs (denoted ei in the model description above, i = A, B, C). These **random** elements in the credit unions' cost functions ensure that a "conglomerate" credit union consisting of the employees of all three firms is not degenerate-i.e., existing with probabilities of either zero or one. Recall that in the first period, all credit unions must consist of a single common bond. The first step in the formation of

Linear City with Three Common Bonds of Occupation
Travel Costs Facing Households Employed by Firms A and B
Demand and Supply Curves for Credit-Union Services
A Case in Which He Credit Union Exists

a credit union is a vote by the potential membership on the management team. Since side payments are allowed, the team that minimizes the sum of the travel costs of all potential members-i.e., which picks the most central location-will win. In a second step, all households decide whether to become members or to purchase **financial** services from a commercial bank.In the second period, multiple-group credit unions are allowed. We iterate through the possible combinations by first allowing mergers between two given credit unions and forcing the third to operate independently (if it exists). Then we allow all three credit unions to merge. In each regime, households vote on the management team (i.e., choose the credit union's location). In particular, households choose between a team that would operate the credit unions independently and a team that would merge them. Because bribing is allowed, the team that maximizes the welfare index over all potential members will win. It is possible that a stand-alone credit union that could not exist on its own becomes part of a multiple-group credit union. The reason is that the post-merger credit union is able to spread its fixed costs over a larger membership. It also is possible that a credit union that could not exist on its own also is not viable as part of a multiple-group credit union. On the other hand, any employee group that is served by a credit union in period I also will be served by a credit union in period 2 because all mergers must be welfare-enhancing. That is, all options for operating credit unions with single common bonds available in period I still are possible after permitting multiple common bonds in period 2.

Table I displays a summary of the simulation results. The table presents two measures of credit-union activity: the fraction of all employee groups served by a credit union and the fraction of households served by a credit union. When only single-employer credit unions are allowed (period 1), only 6 percent of the .30,000 simulated employers (A, B, and C in each of 10,000 simulations) actually sponsor a credit union and only 4 percent of households actually belong to credit unions. Among households that are eligible to join a credit union, some 50 percent do so. All other households use commercial banks to obtain **financial** services. We have chosen parameter values to reflect the fact that single-group credit unions are relatively small and may not be viable for many employee groups.

The bottom part of Table 1 presents results when multiple-group credit unions are allowed (period 2). It is clear that the permissibility of multiple common bonds dramatically increases the viability of credit

unions. This is a general result in the sense that restricting credit union membership to one employee group is a binding constraint, the relaxation of which may increase the beneficial role of credit unions for employees. When two employee groups may be combined in a single credit union (A and B, A and C, or B and Q, the fraction of employee groups in the economy served by a credit union rises to between 14 and 50 percent, while the fraction of households served by a credit union rises to between 4 and 37 percent, depending on the combination. When all three employee groups are allowed to combine in a single credit union (A and B and Q, the fraction of employee groups served by a credit union jumps to 49 percent, although only 30 percent of households are still served.4

Examination of column 6 indicates that multiple-group credit unions comprising groups A and B or A, B, and C are characterized by relatively low participation rates. This reflects the fact that many members of employee groups A and B are located far from any multiple-group credit union, reducing their incentive to join. The credit union formed by employee groups A and C alone, on the other hand -which are located back-to-back-is characterized by a very high participation rate (77 percent of those eligible actually join). This is because the preferences of these two groups overlap.

In general, how likely households are to join credit unions does not depend primarily on whether multiple-group credit unions are allowed (see column 7, where the exception is the credit union comprising groups A and C, the back-to-back case). In other words, participation rates in multiple-group credit unions are not necessarily higher. Rather, it is the fact that more credit unions are viable when multiple common bonds are allowed that is responsible for the expanded role of credit unions in the economy. A comparison of columns I and 3 shows that newly viable multiple-group credit unions are indeed the key to greater credit-union access by households, as the lion's share of all credit unions in every possible configuration in period 2 include multiple common bonds.

The final row of Table I presents the social optimum, which is the welfare-maximizing combination of singleand multiple-group credit unions that is feasible in the economy Multiple-group credit unions serve 93 percent of all employee groups in the social optimum, while single-group credit unions serve only 2 percent. Average household participation rates are similar across the two types of credit unions, with the multiplegroup average slightly higher.

Simulation Results (1)

The averages presented in Table 1 conceal two important features of credit unions in our model, however. Figure 5 is a scatterplot showing the participation rates of all the (optimally formed) credit unions from our 10,000 runs as a function of potential membership. The horizontal scale runs from about 0. 1 (the minimum segment length needed to support a credit union under our baseline parameterization) to 2.0 (the sum of two unit-length segments, corresponding to the maximum potential membership of any multiple-group credit union). The two distinct downwardcurving sets of points represent the declining participation rates of singlegroup (the gray points in the lower curve, ending at 1.0) and multiple-group credit unions (the blue points), respectively. If we were to show each type of credit union plotted in Figure 5 in a separate chart, two important features would be obvious from the average participation rates shown in Table 1.

The first feature is that participation rates of multiple-group credit unions tend to lie above those of single-group credit unions for a given number of potential members. This points to the fact that multiple-group credit unions can be closer to the average member's preferences due to the existence of back-to-back households (i.e., households with different employers but identical preferences for banking services). This effect is due entirely to the households in employee group C in our model, whose preferences overlap those of some households in other employee groups, most importantly group A.

Figure 5 also shows the second important feature of the model that is

not revealed in the table-the downward slope of both main sets of points. Greater potential membership tends to generate lower participation rates. This always holds for single-group credit unions and for multiple-group credit unions that comprise "horizontally neighboring" membership groups only (i.e., groups A and B). Given the travel costs that represent preference heterogeneity among the potential membership, it is not surprising that credit unions that span a more heterogeneous set of households are able to attract proportionately fewer of them. For credit unions that. comprise segments B and C, participation rates initially fall with an increase in the membership base, then rise and later fall again. Credit unions that unite all three employee groups (A&B&C credit unions) exhibit a similar pattern in terms of participation rates. For A&B&C credit unions, a rise in the potential membership is due solely to an increase in the **interval** spanned by group C, which means an increase in the fraction of households located back-to-back. For back-to-back credit unions of the type A&C, the participation rate increases with a growing membership base if (and only if) the overlap of the **intervals** spanned by the two segments increases. If (and only if) the overlap shrinks with an increase in the potential membership, the participation rates shrink, too. To sum up, the overall effect of the size of potential membership on the participation rates depends on the relative importance of the various types of multiple-group credit unions.

Table 2 presents comparative-static results for changes in the parameters t, c, and a (the travel cost parameter, the price of banking services of commercial banks, and the parameter in the multiple-group credit unions' fixed costs, respectively). The first row restates the results of the benchmark simulation summarized in the last row of Table 1. Columns 1-5 show the number of times in the 10,000 runs of the simulation that each configuration of credit unions was optimal. The most frequently preferred configuration was a two-group credit union comprising employee groups A and C (column 3), the back-to-back solution. In this configuration, employees of firm B were sometimes served by a credit union and sometimes not; the feasibility of a credit union for employee group B depends on the size of the membership base and the **random** technology of the potential credit union. The next most frequently preferred configuration involved a three-group credit union. Across all simulations, almost 27 percent of employee groups were left unserved by credit unions even though

Participation Rates as a Function of Potential Membership

all mergers were chosen optimally (this figure is calculated from column 6, which is divided by the total number of employee groups in the simulation, 30,000). It is apparent that participation rates of multiplegroup credit unions (column 10) are dragged down primarily by the relatively low participation rates in the credit unions with horizontally neighboring groups (i.e., credit unions A&B and A&B&C; recall the result from column 6 of Table 1).

The first comparative-static exercise we performed is summarized in the second row of Table 2. When the price of **financial** services offered by commercial banks rises, the fraction of employee groups as well as the fraction of households served by credit unions increases, as expected. From column 6 we know that only 16 percent of employee groups have no credit union after the higher cost of bank-provided services is imposed, while only 46 percent of households use a commercial bank (down from 57 percent in the benchmark case; see column 7). Interestingly, all of the multiple-group credit unions increasingly are preferred when banking services become more costly, while only the singlegroup credit unions become less likely

Simulation Results (2)

to be optimal. A higher price for bankprovided services is predicted by higher banking concentration in the structureconduct-performance paradigm, and our comparative-static result demonstrates that credit unions are indeed likely to benefit from more concentrated banking markets.

The second comparative-static result we computed is summarized in the third row of Table 2. When the cost of travelling to a credit union is increased-intuitively, when preferences for banking services become more idiosyncratic or strongly held-both the fraction of employee groups served by credit unions and the participation rate of households decline (columns 6 and 7-10, respectively). Compared to the benchmark case, the number of single-group credit unions in optimal configurations increases (column 1).

On the other hand, multiple-group credit unions appear somewhat less attractive (columns 2-5).

The last row of Table 2 displays our third comparative-static result. When the fixed costs of production are systematically higher for multiple-group credit unions than for single-group credit unions, the formation of multiple-group credit unions is less advantageous. Relative to the benchmark case displayed in row 1, fewer groups of employees are served by credit unions (column 6). When they exist, credit unions with multiple-group charters have higher participation rates than in the benchmark case (column 10), which is due mainly to the higher representation of pure back-to-back credit unions (column 3). This leaves the overall household participation rate unchanged at the reported two-digit level of precision (column 7).

Taken together, the comparative-static results in Table 2 indicate that the optimal configuration of credit unions in the economy is sensitive to model parameters such as the market price of bank-provided **financial** services, the intensity of preferences for specific bundles of banking services, and the potential extra costs associated with multiple-group charters.

We are now able to state several testable hypotheses that involve the determinants of the participation rate and the average operating costs (i.e., the cost ratio). First, we focus on participation rates at credit unions. Our maintained hypothesis is that a credit union is more successful in providing services to its constituency the less heterogeneous is its membership. This leads to our first testable hypothesis:

HYPOTHESIS 1. A credit union's participation rate falls with the number of its potential members, all else held constant.

Another hypothesis concerns the effects of local banking-market conditions on credit-union participation rates:

HYPOTHESIS 2. A credit union's participation rate rises with the level of concentration in the local banking market, all else held constant.

Next we investigate the validity of our maintained assumption that credit unions face scale economies in production:

HYPOTHESIS 3a. A credit union's cost ratio falls with the number of its potential members, all else held constant.

HYPOTHESIS 3b. A credit union's cost ratio falls with its level of total assets, all else held constant.

Related questions include the effect of multiple-group charters on the cost ratio and the participation rate. Neither the model nor the simulation address the relationship between multiple common bonds and the cost ratio. As for the impact of multiple common bonds on the participation rate, the model and simulation results are ambiguous. Consequently, it is purely an empirical question how multiple-group credit unions affect operating costs and participation rates, holding all else constant.

We examine a subset of all federally chartered and federally insured occupational credit unions in 1996 (see the appendix for details on construction of the dataset and the variables we use). Table 3 provides a breakdown of our sample according to the type of membership group characterizing each credit union. The table distinguishes between credit unions with a single common bond and those with multiple common bonds. Credit unions sponsored by a single educational institution, for example, numbered 299 in our sample. Credit unions with a membership comprising multiple common bonds, most of which were educationally oriented, numbered 469, and so on for the other membership types. Overall, 1,980 credit unions in our sample had a single common bond (41.8 percent of the sample) while

2,753 credit unions had multiple common bonds among the membership (58.2 percent).

In addition to data on individual credit unions, we collected three types of environmental variables. To control for differences in local economic conditions, we gathered levels and computed growth rates of real gross state product for each state. Measures of economic activity may capture systematic differences in demand for credit union services that we have not modeled explicitly We also calculated the Herfindahl index of concentration of bank deposit shares in each credit union's local banking market, since concentration measures often are used to control for differences in the competitiveness of local markets. The index is calculated as the sum of the squared market shares of all participants in each local market. Third, we collected data on population density by county, which might be another factor in credit unions' competition with commercial banks.defined as the number of actual members divided by the number of potential members as specified in the credit union's charter; and 2) COST, the credit union's total operating expenses divided by total assets. There are four independent variables of interest: the number of members (or the number of potential members when we examine participation rates); total assets (for the COST regression); the Herfindahl index of local bank-deposit concentration (HERF); and the indicator variable MULTGROUP, which is equal to one if the credit union has a multiplegroup charter and zero otherwise.

Membership (or potential membership) and total assets are included in the nonparametric part of both regression approaches. They are in logarithmic form and-to avoid simultaneity problems-are lagged by one period. The parametric part of the model includes the other two variables of interest, HERF and MULTGROUP The parametric part also contains the following control variables: the credit union's home state's real gross state product per capita (REALGSPPC) in the PARTICIPATION regression; the log growth rate of real gross state product (GRREALGSP) in the COST regression; and indicator variables corresponding to the credit union's primary field of membership (in both regression approaches). Fields of membership include educational, military, government, manufacturing, and services. Because there is a constant included in the nonparametric part of the regression equation, we must drop one of the membership indicator variables; we chose the educational indicator variable for exclusion.

The variable REALGSPPC in the PARTICIPATION regression controls for preferences for banking services as they may vary with real income. In the COST regression, GRREALGSP serves as a measure for real growth, which is a main factor in the capacity utilization of credit unions.

Distribution of Credit Unions by Type Of Membership (TOM)

In the PARTICIPATION regression, the Herfindahl index is lagged to avoid simultaneity problems that may arise from the interaction between credit-union participation rates and concentration in the local banking market.

Table 4 presents descriptive sample statistics for the dependent and some of the independent variables. The participation rate among sample credit unions ranged from 3 percent to 100 percent, with the median at 62 percent. The median cost ratio was 3.90 percent, with a range of 0.63 to 41.70 percent of assets. Although this range may contain some extreme values, we retain all observations because all of them contain information. In addition, our locally weighted regression approach is somewhat robust to outliers. Total assets ranged from $43,000 to $8.92 billion, with the median credit union holding $6.23 million in assets. The number of actual and potential members ranged from 45 to 1.6 million and 75 to 2.03 million, respectively, while median actual and potential membership counts were 1,865 and 3,198, respectively Finally Herfindahl indexes in relevant banking markets ranged from 0.0535

Descriptive Statistics'

to 1. 00 with a median **value** of 0. 1966 (recall that the index is defined on the **interval** ((0,11).

Our results are presented in two sections corresponding to the dependent variable used. The first section discusses results from regressions using credit unions' participation rate while the second section reports results from regressions using credit unions' cost ratio.

Hypothesis I relates the size of a credit union's potential membership to its participation rate. Regressions including PARTICIPATION use (the lagged **value** of the log of) potential members instead of actual members. Potential members are relevant for evaluating participation rates because all individuals eligible to join constitute the predetermined economic potential that each credit union seeks to exploit.

The series of plots presented in Figures 6-8 are "conditioning plots." The solid lines in Figures 6-8 are point estimates and the dashed lines indicate 90-percent confidence bounds. in each plot, one variable is held at its median **value** while the other variable (identified on the horizontal axis) is allowed to vary. The graph displays the impact of this independent variable on the level of the dependent variable. In other words, the slope of the graph at a particular point reflects the marginal impact of the indepen dent variable at that point. The intercept is not identified in regressions of this type, so only vertical distances are meaningful (not the level itself). In sum, the key to interpreting these graphs is to focus on the slope of the curve and on the vertical differences moving along the horizontal axis .5

Participation Rate - Number of Potential Members

Cost Ratio - Number of Members

Cost Ratio - Total Assets

Figure 6 provides evidence supporting Hypothesis I (a negative relationship between participation rates and potential membership). The plot supports Hypothesis 1 because the lower confidence bound at small credit unions (small number of potential members) lies above the upper confidence bound for large credit unions, except for the very smallest and very largest credit unions, where the small number of observations widens the confidence **intervals**. This confirms our findings when we simulated the theoretical model (recall Figure 5) and is consistent with the idea that larger membership pools contain greater heterogeneity of preferences for banking services. This leads to greater differences in the most preferred bundle of banking services between the median member and members in the tails of the preference distribution.

An existing single-group credit union that adds one or more membership groups to its common bond encounters both benefits and costs of expansion. On the one hand, adding membership groups whose preferences are close to those in the existing field of membership (the back-to-back case) may increase the participation rate. This outcome would be predicted as a result of a reduction in average operating costs and hence in the credit union's price of banking services. On the other hand, for a given credit-union charter, a higher number of potential members typically means more heterogeneity and thus a lower participation rate as the travel distance of the members located in the tails of the preference distribution increases.

Table 5 presents results from the parametric part of the model. The results provide evidence on the question of whether credit-union participation rates differ when comparing credit unions with a single common bond to those with multiple common bonds, holding all else equal. Recall from Tables I and 2 that credit unions with multiple-group charters will (on average) have higher participation rates than single-group credit unions if back-to-back membership combinations dominate (i. e., if there is a significant overlap of banking preferences among employees of different firms). Table 5 indicates that multiple-group credit unions in our sample indeed have higher participation rates, perhaps reflecting the ability of multiple-group credit unions to capitalize on similar preferences among employees of different firms.

Another interesting result in Table 5 is the positive and significant coefficient on the lagged Herfindahl index of bank deposit concentration in

credit unions' local markets (Hypothesis 2). This indicates that, the more concentrated its local banking market is, the higher a credit union's participation rate will be. In other words, credit unions may provide an attractive alternative for consumers who face a relatively uncompetitive local banking market.

Hypotheses 3a and 3b refer to tests of an important maintained assumption of our model, namely, that a credit union's operating expenses should decline with an increase in its scale of operation. We also would like to know whether serving multiple-group memberships is costly Figure 7 plots COST against the number of members, while holding the level of total assets constant at its median **value** in the sample. The plot shows that COST decreases sharply beyond a certain threshold level of membership. For small credit unions, average costs seem to increase slightly with the number of members. While wide confidence **intervals** indicate that in this range the relationship between average costs and number of members is estimated imprecisely, an initial increase in the average operating costs actually might be supported by the data. For small credit unions, subsidies (such as rent-free office space, volunteer workers, etc.) tend to be relatively more important than for large units. As these subsidies become less important for credit unions with higher numbers of members, measured operating costs might approach shadow operating costs. Overall, the findings support our maintained assumption of declining average costs as the scale of operations increases, Similar evidence is provided in Figure 8, which is generated by the same regression that produced Figure 7. Figure 8 plots the influence of another measure of the size of operations, total assets, on the credit union's average operating costs (Hypothesis 3b) while holding the numbei of members constant at its median **value** in the sample.

Table 6 presents results from the parametric part of the model. The results indicate that there is a positive relationship between the existence of multiple membership groups in a credit union and COST. One might think that multiple-group credit unions would have high cost ratios due

Participation Rate

to agency costs. According to this line of reasoning, as membership groups try to free-ride on each other's monitoring, supervision of management might be inefficiently low. As Emmons and Schmid (1999) show, however, there is no evidence of multiplegroup charters causing agency costs.

Finally, the significantly negative coefficient on the Herfindahl index in Table 6 implies that higher levels of bank concentration in a local market lead to lower levels of the cost ratio reported by credit unions. One possible explanation is that less intense competition from banks allows credit-union managers to enjoy a "quiet life." For example, credit unions may be able to attract or retain members with lower marketing efforts or lower quality services than would be the case in a more competitive market. On the other hand, the quiet life that comes with less competitive markets might allow greater scope for managerial agency costs. If the latter were the case, however, it would generate predictions opposite to our empirical findings of lower average operating expenses.

Continued expansion of credit unions has been accompanied by public debate, and courtroom confrontations. Advocates argue that credit unions provide needed competition to banks and thrifts in local markets for retail **financial** services. Opponents, including most notably the banks and thrifts themselves, point to various subsidies to credit unions that create an unlevel playing field. Previous research findings do not provide unambiguous conclusions favorable to either camp, while recent federal legislation favorable to creditunion expansion merely has intensified the debate. More research into the fundamental operation of credit unions is needed.

In this article, we investigate the relationships between several features of credit unions, namely the number of members in a credit union,

the amount of total assets on its balance sheet, and the existence of single and multiple common bonds among its membership, on the one hand, and two measures of credit-union effectiveness, on the other. We also examine the effect of several environmental variables, including economic conditions and banking concentration in the local market, on credit-union operations.

We find that a larger potential creditunion membership translates into lower credit-union participation rates. Credit unions with multiple common bonds, holding all else constant, have higher participation rates. We also find evidence that credit unions in more concentrated banking markets exhibit higher participation rates.

While greater asset size appears to be associated with lower average operating costs, holding all else equal, we find that a larger number of members is associated with a lower cost ratio only for larger credit unions. Thus, asset size and the size of the membership are distinct aspects of creditunion operations. Multiple-group credit unions have higher costs on average, all else equal. We also find that credit-union cost ratios are lower in more concentrated banking markets, perhaps indicating that credit unions can economize on marketing or service provision when competition from banks is less intense.

Our findings are particularly interesting in light of the recent AT&T Family Credit Union case decided by the Supreme Court in February 1998, and its sequel in the U.S. Congress that culminated in the Credit Union Membership Access Act of August 1998. This new federal legislation upholds the right of federally chartered credit unions to grow under an expansive definition of the common-bond requirement. The new law allows multiple groups of members to belong to a single credit union as long as the members of each group are united by a common bond. This statute therefore upholds regulatory actions taken in recent years and overturns the Supreme Court's narrow reading of the 1934 Federal Credit Union Act restricting a federal credit union to a single common bond.

We analyze a dataset comprising all federally chartered and federally insured credit unions during the year 1996. The dataset was obtained from the Report of Condition and Income for Credit Unions (NCUA 5300, 5300S), produced by the National Credit Union Administration (NCUA). These reports are issued semiannually in June and December. We used

Cost Ratio

the December data. The flows in the December income statements include the entire year of 1996.

We concentrate on the following Types Of Membership (TOM) groups among occupationally based credit unions: educational; military; federal, state, and local government; manufacturing; and services. This means that we do not include community credit unions, associational credit unions, or corporate credit unions. Lists of TOM classification codes are from the NCUA (Instruction No. 6010.2, July 28,1995). We excluded observations for any of the following reasons:

Missing TOM codes.

Activity codes other than "active.'

Number of members or of potential members not greater than one; applies to actual and to lagged values.

Nonpositive values for total assets or lagged total assets.

Zero number of employees.

Zero **value** for "employee compensation and benefits."

Total assets, number of members, potential number of members, and the Herfindahl index were all lagged one year (i.e., 1995 values). All other observations are from year-end 1996.

We calculated county-specific Herfindahl indexes as measures of concentration of the local banking market. A Herfindahl index is defined as the sum of squared market shares. We measured market shares by the fraction of total bank deposits (as of June 30) within a county based on FDIC Summary of Deposits data. These data are available online at <http://www2.fdic.gov/sod/>.

We used either the log level of Real Gross State Product (REALGSP) or its log growth rate to control for cross-sectional differences in macroeconomic conditions facing credit unions. The REALGSP data are in millions of chained 1992 dollars. We obtained the data from the U.S. Department of Commerce, Bureau of Economic **Analysis**, Regional Economic **Analysis** Division. The data are available online at <http://wwwbea. doc.gov/bea/regional/data.htm>.

Population density at the county level was calculated by dividing the total county population by the total land area of the county (in square miles). Both the county population and land area data were obtained from the U.S. Census Bureau <http://www.census.gov>. The population data are Census Bureau estimates as of July 1, 1996. The land area measurements are from the 1990 census.

We transformed the dependent variables in some cases to ensure that they are not bounded. These transformations are necessitated by the assumption of normally distributed error terms. For variables that are restricted to the positive orthant of real numbers, we substitute their natural logarithms. For variables expressed as fractions (i.e., restricted to the **interval** [0, 1]), we applied the logit transformation log(y10-y)), with log being the natural logarithm. In this case, observations equal to one were eliminated from the set of observations; there were no cases in which the transformed variable equaled zero.

Definitions of variables and underlying data sources are listed below. For data taken from the Report of Condition and Income for Credit Unions, produced by the National Credit Union Administration, the relevant item numbers are in brackets.

Dependent Variables. We employed two dependent variables in the regressions:

1) Participation Rate (PARTICIPATION): Number of actual credit-union members [CUSA60911 divided by the number of potential members [CUSA60921. In the regressions, we use the logit transformation log(yI(I-y)). No zero values for the number of members occurred. Fortytwo cases of full participation (y = 1) were eliminated from the dataset for these regressions only.

2) Cost Ratio (COST): Total operating expenses [CUSA41301 divided by total assets [CUSA21701. In the regression, we use log values.

Independent Variables. When total assets (measured in units of one dollar), the number of members, or the number of potential members served as regressors, they were lagged by one period and transformed into natural logarithms.

1. MULTGROUP: equal to one if the credit union has multiple groups; zero otherwise.

2. HERE Sum of squared market shares of commercial banks within a county based on total bank deposits. By definition, the Herfindahl index is greater than zero; its maximum **value** is one.

3. REALGSPPC: Real gross state product per capita (chained 1992 dollars).

4. GRREALGSP: Logarithmic changes in the real gross state product (chained 1992 dollars).

5. POPDENS: Population Density, people per square mile in each local banking market.

6. TOM code variables: equal to one if the credit union is of a specific type (educational, military, government, manufacturing, or services). Because we use an intercept in (the nonparametric part of) the regression, the TOM code variable for the educational credit union was dropped.

THE CREDIT UNION MEMBERSHIP ACCESS ACT

President Clinton signed the Credit Union Membership Access Act on August 7, 1998, following approval in the Senate on July 28 and in the House of Representatives on August 4. The act substantially reverses a Supreme Court ruling handed down on February 25, 1998, that would have barred federally chartered credit unions from accepting multiple membership

groups, each with its own common bond.

This landmark credit-union legislation represents a major defeat for the top lobbying group representing commercial banks, which had argued successfully at the Supreme Court that credit unions with multiple common bonds violated both the letter and the spirit of federal legislation dating, from 1934. The subsequent legislative response in support of multiple common bonds at credit unions was swift and overwhelming, passing both chambers with large majorities.

The act contains three provisions upholding the rights of federal credit unions to serve membership groups encompassing multiple common bonds. First, all federal credit unions that already included multiple common bonds before February 25, 1998, were allowed to continue operating without interruption. Second, all federal credit unions were given the right to accept additional membership groups with multiple common bonds so long as the relevant groups have fewer than 3,000 members. Third, the act gives the National Credit Union Administration the right to grant exemptions to the 3,000-member limit under certain circumstances, such as when the group in question could not reasonably support its own credit union.

The act also:

Requires annual independent audits for insured credit unions with total assets of $500 million or more.

Authorizes and clarifies a federally insured credit union's right to convert to a mutual savings bank or savings association without prior NCUA approval.

Limits business loans to members to 12.25 percent of total assets.

Establishes new capital standards for insured credit unions similar to those enacted for banks and thrifts in 1991.

Gives the NCUA authority to base deposit-insurance premiums on the reserve ratio of the insurance fund.

Directs the Treasury to report to Congress on differences between credit unions and other federally insured **financial** institutions, including the potential effects of applying federal laws-including tax laws-to credit unions.

Hailing the new legislation, President Clinton said, "This bill ensures that consumers continue to have a broad array of choices in **financial** services .... and [makes] it easier for credit unions to expand where appropriate." Meanwhile, a spokeswoman for the American Bankers Association termed it "ironic" that the bill was presented as a measure to protect credit unions because in the long run, she said, it will dilute them, turning them into larger and larger institutions.

Source: BNA Banking Report, "House Passes Credit Union Bill; Clinton Wastes No Time Signing It," August 10, 1998, Vol. 71, No. 6.

American Bankers Association. The Credit Union Industry: Trends, Structure, and Competitiveness (Washington, D.C., 1989).

BNA Banking Report. "House Passes Credit Union Bill; Clinton Wastes No Time Signing It," August 10, 1998, Vol. 71, No. 6. Oevelard, William S., and Susan 1. Devlin. "Locally Weighted Regression: An Approach to Regression **Analysis** by Local Fitting," Journal of the American StafiscalAssodaion (September 1988), pp. 596-610.

Emmons, William R., and Frank A. Schmid. "Wages and Risk-Taking in Occurational Credit Unions: Theory and Evidence," this Review (March/April 1999), pp.13-31.

Hotelling, Harold. "Stability in Competition," Economic Journal, (Mor(h 1929), pp. 41-57.

Kane, Edward J., and Robert Hendershott. "The Federal Deposit Insurance Fund that Didn't Put a Bite on U.S. Taxpayers," Journal of Ba,king and **Finance** (September 1996), pp. 1305-27.

Milgrom, Paul, and John Roberts. Economics, Organization and Management, Prentice Hall, 1992.

Supreme Court. "National Credit Union Administration, Petitioner, v. First National Bank & Trust Co., et al.; AT&T Family Federal Credit

Union, et al., Petitioners, v. First National Bank and Trust Co., et al." Decided Feb. 25, 1998. Nos. 96-843, 96-847. 118 S. Ct. 927.

U. S. Treasury Department. Credit Unions, U.S. Government Printing Office, 1997.

I We concentrate on federally chartered credit unions because the NCUA does not vouch for the o(curacy of data provided by state-chartered credit unions, which report directly to their state's regulatory authorities.

I The estimated 70 million current credit-union members represent a bit more than 34 percent of the 1996 U.S. population over 16 years of age numbering 204 million (U.S. Census Bureau, <http://www.census.gov>).

3 Free-riding is when members choose not to exert monitoring effort because they assume someone else will do it for them.

Household segments not involved in a merger during the second-period simulations face the same economic situation as in the first period. Consequently. they come to the some decision during the second period of whether to operate a credit union as they did during the first period.

William R. Emmons is a research economist and Frank A. Schmid is a senior research economist at the Federal Reserve Bank of St. Louis. Robert Webb and Marcelo Williams provided research assistance.

Special Features: Formula; Chart; Graph; Table
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 8110 (CN=Commercial banking); 1130 (CN=Economic theory)
Descriptors: Studies; Credit unions; Statistical **analysis**; Memberships; Simulation; Economic models
Geographic Names: United States; US
PRINT MEDIA ID: 23954

36/9/3 (Item 3 from file: 268)
00371853
Birth, growth, and life or death of newly chartered banks
DeYoung, Robert
Economic Perspectives (Federal Reserve Bank of Chicago) , v 23 , n 3 , p 18-35 , Third Quarter 1999
Document Type: Journal Article **ISSN:** 1048-115X **Journal Code:** ECP **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 09847

Abstract:

Thousands of new commercial banks have been chartered in the US over the past 2 decades. The way in which the **financial** characteristics of new banks evolve over time is documented. A simple theory of why and when new banks fail is developed, and the theory is tested using a variety of methods. The results suggest that the primary determinant of new bank failure is how new the bank is. Brand new, but unprofitable, banks are typically protected from failure by large initial capital cushions. However, equity cushions at de novo banks typically decline to established bank levels several years before their earnings become stable enough to justify these relatively low levels of capital. The implications of this result for capital regualtion of newly chartered banks are discussed.

Text:

Introduction and summary

Thousands of new commercial banks have been chartered in the U.S. over the past two decades. As the U.S. banking industry continues to consolidate, these de novo banks are potentially important for preserving competition and providing credit in local markets. However, like other new business ventures, newly chartered banks initially struggle to earn profits, and this **financial**amp; fragility makes them especially prone to failure. In this article, I document the **financial**amp; evolution of the typical de novo bank and develop and test a simple theory of why and when new banks fail.

Recent decades have seen an upsurge in the number of mergers and failures among new banks. Figure 1, panel A shows the annual change in the number of commercial bank charters in the U.S. since 1966. Prior to 1980, the reduction in bank charters due to mergers and failures was relatively stable at about 100 charters per year, or about 1 percent of the industry total (figure 1, panel B). The pace accelerated greatly after 1980, and since 1986 about 600 charters, or 5 percent to 6 percent of the industry total, have disappeared each year due to mergers and failures.

To a large extent, this tremendous consolidation can be explained by the repeal of federal and state laws that restricted branch banking and interstate banking. As these restrictions gradually were relaxed, banking companies expanded their geographic reach by acquiring thousands of other banks, and reduced their overhead expenses by converting thousands of affiliate banks into branch offices. This geographic expansion, combined with newly deregulated deposit rates, increased competition between commercial banks just when new information technology was allowing mutual funds, insurance companies, and the commercial paper market to compete for banks' traditional loan and deposit businesses. Under these new competitive conditions, many commercial banks became more vulnerable to economic downturns, and thousands of banks failed during the 1980s and early 1990s. Over the past two decades, the combined effect of these mergers and failures has reduced the number of commercial banks in the U.S. by nearly 40 percent.

This consolidation has been partially offset by a recurring wave of new bank charters. As shown in figure 1, panel A, over 3,000 de novo commercial banks have been chartered by state and federal banking authorities since 1980. It is generally believed that these newly chartered banks can help restore competition in local markets that have experienced a large amount of consolidation. It is also commonly believed that these newly chartered banks can help replace credit relationships for small businesses whose banks failed or were acquired or reorganized. However, before a newly chartered bank can provide strong competition for established banks and before it can be a dependable source of credit for small businesses, it must survive long enough to become financially viable.

I begin by examining the conditions under which investors are likely to start up new banks, including the influence of business cycles, merger activity in local banking markets, and the policies of federal and state chartering authorities. Next, I track the evolution of profits, growth rates, capital levels, asset quality, overhead costs, and funding mix at more than 1,500 commercial banks chartered between 1980 and 1994. These data suggest that newly chartered banks pass through a period of **financial**amp; fragility during which they are more vulnerable to failure than established banks. Specifically, new bank capital ratios quickly decline to established bank levels, but new bank profits improve more slowly over time before attaining established bank levels.

(Table Omitted)

Captioned as: FIGURE 1

Based on these empirical observations, I develop a simple life-cycle theory of de novo bank failure, in which the probability of failure at first rises, and then declines with the age of the new bank. I use hazard function **analysis**amp; to test this simple theory for 303 new commercial banks chartered in 1985, just as the wave of bank failures shown in figure 1 was picking up steam. The tests offer support for the simple theory. On average, the results suggest that newly chartered banks are less likely to fail than established banks during the first few years of their lives; however, new banks quickly become substantially more likely to fail than established banks; and, over an extended period of time, new bank failure rates gradually converge to the failure rates of established banks.

What are the implications of these results for bank supervision and bank competition policy? The results suggest that the policies in place during the 1980s successfully insulated new banks from economic disruptions early in their lives, but were less successful in preventing new banks from failing after the initial years. Clearly, de novo bank failure rates could be reduced by requiring investors to supply higher amounts of start-up capital or by requiring banks to maintain extranormal capital-to-asset ratios in the early years-indeed, the latter policy option was adopted by federal bank supervisors during the 1990s. However, failure-proofing de novo banks is not an optimal policy. The social costs of small bank failure are relatively low, and setting higher capital requirements would at some point discourage investment in new banks and thereby limit the competitive benefits of de novo entry.

Birth of new banks

As illustrated in figure 1, panel A, the number of new banks started up each year has ebbed and flowed over the past three decades. There are a number of explanations for these patterns. Like all new business ventures, new banks are more likely to form when business conditions are good. For example, new bank charters bulged to well over 250 per year during the general economic expansion of the mid-1980s. This high rate of bank start-ups also coincided with the relaxation of unit banking laws in a number of states, laws that had prevented banking companies from operating affiliates in multiple locations. The steady decline in new charters during the late 1980s and early 1990s, which bottomed out at about 50 new banks per year, also had multiple causes. Difficult times in regional banking markets made new bank start-ups unprofitable in many regions (bank failures reached their peak in 1988), and a national recession in the early 1990s reinforced this trend. New bank charters have been on the increase since then, reaching over 100 per year in 1997 and 1998, in large part due to the extended economic expansion of the 1990s.

Conditions in local banking markets also influence bank start-ups. Moore and Skelton (1998) find that there are more de novo banks 1) in markets that are experiencing healthy economic growth, 2) in highly concentrated banking markets in which competition among existing banks is weak, and 3) in markets where small banks are under-represented and, hence, small businesses are not being adequately served. These results imply that new banks will be more likely to start up in local markets where mergers have reduced the number of competing banks, and where the resulting market power has reduced the level of banking services. In such markets, new banks should receive a profitable welcome from customers unhappy with paying high prices for **financial** services or from businesses whose credit relationships were disrupted when their bank was acquired or failed. Researchers only recently began investigating these phenomena, so there is not yet a consensus on the results. In a study of de novo bank entry in all U.S. markets between 1980 and 1998, Berger, Bonime, Goldberg, and White (1999) find that the probability of de novo entry is higher in local markets that have experienced mergers or acquisitions during the previous three years, particularly mergers and acquisitions involving large banking organizations. In contrast, Seelig and Critchfield (1999) find that local market entry by acquisition deters entry by de novo banks and thrifts.'

Their results are based on a study of de novo banks and thrifts between 1995 and 1997, a time when banking conditions were exceptional and restrictions on geographic mobility were virtually nonexistent.

Differences in the policies of the legal authorities that grant commercial bank charters can also affect the rate and location of new bank start-ups. A de novo national bank receives its charter from the Office of the Comptroller of the Currency (OCC), while a de novo state bank receives its charter from the banking commission of the home state. The OCC has **historically** been more liberal in granting charters than most state authorities. Its policy has been that market forces, not the chartering authority, should determine which local markets need and can support new commercial banks. In contrast, many state chartering authorities have **historically** applied convenience and needs tests when considering applications for new bank charters, denying applications if they judge that the convenience and needs of the banking public are already adequately served. Although this federal-state difference in chartering philosophy has diminished over time, DeYoung and Hasan (1998) find that national banks were chartered with greater frequency than state banks during the 1980s and early 1990s, and that the **financial** performance of de novo national banks initially lagged that of de novo state chartered banks.2 This suggests that national banks chartered during the 1980s were likely to have had a higher probability of failure than newly chartered state banks operating under similar economic and market conditions.

A concern shared by all chartering authorities is that newly chartered banks start out with enough equity capital to survive through the several years of negative earnings and rapid asset growth that is typical of de novo banks. The dollar amount of startup **financial** capital required for approval might be $3 million, $10 million, or even as much as $20 million, depending on the proposed location and business plan of the prospective bank. Larger amounts of start-up capital are generally required for urban banks, for banks locating in vibrant economic markets, and for banks with business strategies that feature fast growth (for example, a new Internet bank).

Once a new bank opens its doors for business, regulatory scrutiny shifts from the applications staff to the examination staff. Bank supervisors pay closer attention to newly chartered banks than to similarly situated established banks, although the difference in treatment varies depending on the new bank's primary regulator. Federal Reserve supervisors will conduct full scope examinations for safety and soundness at a newly chartered bank at six-month **intervals** (established banks are examined every 12 to 18 months) and will continue to schedule exams at this frequency until the bank receives a strong composite CAMEL rating (that is, a rating of 1 or 2) in two consecutive exams. The Federal Deposit Insurance Corporation requires that all newly chartered state and national banks maintain an 8 percent tier 1 equity capital-to-riskbased assets ratio for their first three years of operation, while the Federal Reserve requires new state chartered Fed member banks to hold this ratio above 9 percent for three years. These temporary extranormal capital requirements for new banks (the tier 1 requirement for established banks to be considered adequately capitalized is only 4 percent) are a relatively recent supervisory response to de novo failure experience of the 1980s. Bank supervisors also prohibit new banks from paying out dividends for several years and, in some cases, require new banks to maintain minimum levels of loan loss reserves.

Evolution of new banks

Relatively few research studies have examined how banks grow and evolve in the years immediately after they receive their charters.3 Brislin and Santomero (1991) show that the **financial** statements of a new bank can fluctuate rapidly and dramatically during its first year. A handful of studies have examined how the profitability of de novo banks grows over time (for example, Hunter and Srinivasan,1990, and DeYoung and Hasan, 1998). Another strand of research documents how small business

lending becomes less important to de novo banks as they mature (for example, DeYoung, Goldberg, and White,1999). In this section, I analyze how a broad group of de novo bank characteristics not typically considered in the literature evolve over time, including de novo bank profits, growth rates, capital ratios, sources of income, **financing**lt; mix, overhead ratios, and loan quality.

Each of the eight panels in figure 2 examines a different **financial**lt; ratio and compares its average **value**lt; for a sample of de novo commercial banks to its average **value**lt; for a sample of established commercial banks. The de novo bank sample includes 4,305 observations of commercial banks that were chartered between 1980 and 1994, were between one and 14 years old when they were observed, and were located in urban banking markets. The established bank sample includes 4,305 observations of commercial banks that were at least 14 years old when they were observed, operated in urban banking markets, and were similar to the de novo banks in terms of asset size. These two samples of banks were originally constructed by DeYoung and Hasan (1998). Box 1 contains additional details about the two bank samples.

To construct each of the graphs in figure 2, I divided the de novo banks into 14 separate age groups (one-year old banks, two-year old banks, etc.). I then calculated the median average for the **financial**lt; ratio in question-say, return on assets (ROA)--for each age group. Plotting these 14 average values in chronological order creates a time path showing how ROA evolves as the typical de novo bank matures. Finally, I superimposed the **value**lt; of ROA at the 25th, 50th, and 75th percentiles of the established bank sample as horizontal lines over the de novo bank time path. These horizontal lines serve as maturity benchmarks against which to compare the progress of de novo banks over time. The rate at which the de novo time path converges with the maturity benchmarks indicates the speed at which the de novo banks mature.

While each of the individual graphs in figure 2 has a straightforward interpretation when considered in isolation, these eight panels reveal a richer story when they are interpreted in conjunction with each other. For example, by itself the return on assets (ROA) graph (panel A) merely confirms the results of existing studies of de novo bank profitability, that is, that the typical new bank loses money until it is about 18 months old and continues to underperform the average established bank for about a decade. But when the ROA graph is considered together with the assetgrowth (panel B) and equity-to-asset (panel C) graphs, a simple theory of de novo failure begins to emerge. De novo banks average an extraordinary 20 percent annual rate of growth during the first three years of their lives. While this fast growth rate is increasing the amount of assets against which new banks need to hold equity capital, the losses suffered during the first and second years of these banks' lives are depleting their equity capital. Despite initially high capital levels, the equity-to-asset ratio of the typical new bank declines very quickly, entering the established bank range after just three years. Thus, panels A, B, and C suggest the probability of failure should increase as new banks pass their third year of life-- their capital has declined to established bank levels by year three, but their asset growth and profitability do not converge with those of established banks until at least year ten.

The remaining five panels in figure 2 are consistent with the simple theory of de novo bank failure suggested by the ROA, asset growth, and equity-toasset panels. For example, newly chartered banks initially have almost no nonperforming loans (panel D). This is because these banks' loan portfolios are composed disproportionately of unseasoned loans made recently to borrowers who demonstrated strong **financial**lt; fundamentals. However, as time passes some of these new borrowers will naturally run into trouble, and the quality of de novo banks' loan portfolios will naturally decline. This happens quite quickly for the typical de novo bank, as its level of nonperforming loans rises slightly above the median level for established banks after three years-just as de

novo banks are depleting their excess capital cushions and well before new bank profitability rates have matured.

The slow rate at which de novo bank profitability improves appears to be attributable more to cost factors than to revenue factors. Although the percentage of de novo bank assets invested in interest-bearing assets, such as loans and securities, starts out relatively low and increases only slowly over time (panel E), the typical de novo bank outperforms one-quarter of the established banks in this area after only three years. (De novo bank ROA does not reach the 25th percentile benchmark until year six.) Even more impressive is the speed at which new banks develop the ability to generate fee income (panel F). The typical de novo bank outstrips the average established bank in feebased revenues after only three years, and outperforms three-quarters of the older banks in this area after about nine years. By virtue of their newness, de novo banks may be less constrained by the inertia of existing customer relationships and existing employee habits and, therefore, may be better able to impose fees on retail customers or to enter into less traditional fee-generating lines of business. In addition, de novo banks tend to start up in markets where business conditions are strong, and selling fee-based **financial**lt; services may be easier in these markets.

In contrast to their reasonably strong ability to generate revenue, newly chartered banks have a diffcult time controlling expenses. De novo banks initially use large deposits twice as intensively as do established banks, and this disparity only slowly disappears (panel G). This suggests that de novo banks tend to **finance**lt; their fast asset growth by purchasing funds rather than by growing their core deposit base. All else being equal, this is an expensive and potentially risky **financing**lt; strategy, because large depositors are more sensitive to changes in interest rates than are retail depositors and require higher rates to leave their funds in the bank. The accounting efficiency ratio graph (panel H) indicates that newly chartered banks also have relatively high levels of overhead expenses (for example, branch locations, labor expenses, and computer equipment) and that these fixed factors of production are not used at near full capacity for a number of years. Excess overhead capacity not only depresses bank profitability but, by increasing operating leverage, it makes bank profits more sensitive to fluctuations in bank revenues.

Note that each of the panels in figure 2 exhibits what is known as survivor bias, because some de novo banks fail before they are 14 years old. For example, average de novo ROA equals approximately 0.4 percent for the three-year old banks, which is about twice as large as the average ROA of 0.2 percent for the twoyear old banks. For the most part, this substantial improvement can be attributed to better performance as young banks grow older and larger. But some amount of this improvement occurs because some of the most unprofitable de novo banks failed between years two and three and dropped out of the sample. Although this second explanation is responsible for only a small amount of the large increase in ROA (as we shall see, very few de novo banks fail after only two years of operation), it is a good illustration of how survivor bias can affect our results. Thus, the most exact way to interpret the ROA time path is as follows: If a newly chartered bank survives to be three years old, one would expect its ROA to be about 0.4 percent. I revisit the issue of survivor bias when I estimate time to failure models in a later section (see Estimating hazard functions section, starting on page 26).

Hypothetical hazard rates

The time paths in figure 2 imply that de novo banks will at first be very unlikely to fail, perhaps even less likely to fail than established banks. Despite the losses typically incurred during their first year of operation, de novo banks initially have very high cushions of equity capital and very low levels of nonperforming loans. But the time paths in figure 2 also imply that de novo banks become dramatically more likely to fail as time passes, and quickly may become more likely to fail than

established banks. As de novo banks age, their initially high capital cushions and low nonperforming loan ratios move rapidly toward established bank levels-much more rapidly than their profitability reaches established bank levels.

The combined effect of these **financial** ratio time paths on the timing and probability of de novo bank failure is suggested by the hypothetical hazard functions in figure 3. A hazard function tracks changes over time in the hazard rate, which is simply the probability that a bank will fail at a particular time, given that it has survived through all of the previous periods leading up to that time.4 The horizontal line at P* represents the hypothetical hazard rate for established banks, and the curved line plots the hypothetical hazard rate for newly chartered banks. Although this figure is highly stylized, the relative shapes of the two functions are consistent with the combined **financial** ratio time paths shown in figure 2.

The constant, non-zero hazard rate depicted in figure 3 for established banks is an obvious simplification. **Historically**, established banks are more prone to failure during recessionary periods, and almost completely unlikely to fail during expansionary periods. This simplification focuses attention on the issue of primary interest here, the failure rate of newly chartered banks relative to the failure rate of established banks.

(Graph Omitted)

Captioned as: FIGURE 2

The hypothetical hazard rate for newly chartered banks starts out at zero in figure 3, which makes sense because these banks are so heavily capitalized at the outset. But, as we saw in figure 2, de novo bank capital ratios decline to established bank averages after about three years, while de novo bank profits, asset quality, and growth rates do not reach (or return to) established bank levels for around ten years. When these time paths are considered simultaneously, they imply the hypothetical patterns displayed in figure 3. The hypothetical hazard rate for new banks increases at first (for example, between ages one and three) as new banks become increasingly vulnerable to economic fluctuations; it exceeds the established bank hazard rate for a time (for example, after year three); and it eventually declines to converge with established bank levels (for example, around year ten). Regardless of the exact shape and timing of the de novo bank hazard function, it must eventually converge with the established bank hazard function, because by definition new banks that survive eventually turn into established banks.

(Graph Omitted)

Captioned as: FIGURE 3

A rough way to check the relative accuracy of the hypothetical hazard functions drawn in figure 3 is to calculate Z-score probabilities of failure for de novo banks and established banks. The Z-scores are constructed as follows: (Formula Omitted)

The Z-score indicates the number of standard deviations that ROA would have to fall below its average **value** in order to wipe out 100 percent of the bank's equity capital. For example, if a bank has 5 percent equity capital and, on average, it earns ROA of 1 percent with a standard deviation of 1 percent, then its Z-score would equal 6.00. In this case, the bank's ROA would have to decline by 6 standard deviations below its average (to -5 percent) for its losses to exhaust its capital cushion. Thus, the higher a bank's Z-score, the lower its probability of failure. Z will increase (that is, the probability of failure will decrease) with higher levels of average ROA; Z will increase with higher levels of equity to assets; and Z will increase with lower variability in ROA.5

Table 1 displays Z-scores for the established bank sample, for the de novo bank sample, and for several subsamples of de novo banks. All of these calculations employ the data used to construct the graphs in figure 2. For each sample or subsample of banks, Z is calculated using the median average of ROA, the mean average of equity/assets, and the crosssectional standard

deviation of ROA. (I use the median ROA because the mean ROA is skewed downward by banks that incurred large losses.) Because these Z-scores are averages, they represent the likelihood of failure for the typical bank in each sample.

In general, the calculations shown in table I suggest that becoming insolvent is a relatively unlikely event for the typical bank in these samples. For example, the lowest Z-score (highest probability of insolvency) is 3.01, or about 3 standard deviations, for the average three- to five-year old bank. Assuming that Z is normally distributed, this implies only a 13 in 1,000 (0.13 percent) chance of becoming insolvent. Given the large number of bank failures during the sample period (see figure 1, panel A), the level of the failure probabilities implied by these Z-scores is probably too low.6 However, these Z-scores are still useful, because they summarize the information in figure 2 into a single number that ranks the probability of failure across banks of different ages.

(Table Omitted)

Captioned as: TABLE 1

Overall, the **analysis** suggests that newly chartered banks are more likely to fail than established banks: The average de novo Z-score of 3.97 is considerably smaller than the average established bank Z-score of 7.48. On average, de novo banks and established banks have nearly identical capital-to-asset ratios, so any difference in their implied failure rates must be due to the level and variability of ROA. Indeed, the median de novo bank ROA is only about half as large as the median established bank ROA (.0057 versus .0097), and ROA is nearly twice as variable across the de novo banks than across the established banks (.0230 versus .0129).

Analyzing the Z-scores across de novo banks of different ages provides some support for the shape of the de novo bank hazard function in figure 3. The implied probability of failure is relatively low for banks less than three years old (Z = 6.00); is substantially higher for three- to five-year old banks (Z = 3.01); and then gradually declines toward established bank levels for banks that survive beyond five years (Z = 4.28) and beyond ten years (Z= 5.70). Looking at the components of these average Z-scores reveals why the probability of failure changes as new banks mature. The youngest group of de novo banks are the least likely to fail because their earnings are relatively stable (although they average near zero) and their capital cushions are large. The three- to five-year old de novo banks are more likely to fail because, although they have higher average earnings, their capital cushions have been depleted and their earnings are highly variable. Once banks are five to ten years old, increasing earnings, increasing capital, and declining earnings volatility all contribute to a reduced probability of bank failure.

Estimating hazard functions

Next, I test whether the hypothetical hazard functions in figure 3 accurately depict the relative rates at which newly chartered banks and established banks fail. The Z-score **analysis** discussed above provides some support for these hypothetical hazard functions, but that evidence is crude at best and suffers from survivor bias in the data. In this section, I employ more sophisticated techniques to estimate hazard functions for both newly chartered and established banks. These techniques explicitly account for survivor bias caused by failures and acquisitions during the sample period. In addition, these techniques generate continuous (or nearly continuous) hazard functions that can be plotted against time, making them easy to compare with the shape of the hypothetical hazard functions in figure 3. Finally, one of these techniques tests whether differences in de novo and established bank failure rates are caused by differences in these banks' locational, regulatory, or organizational characteristics.

Data on bank failures

Table 2 displays some summary statistics for a bank failure dataset Federal Reserve Bank of Chicago staff created for the purpose of this

study. This dataset contains 56 quarters of information on 2,653 banks from 1985 through 1998, and is constructed from the "Reports of income and condition" (call reports) and from the failures, transformations, and attributes tables in the National Information Center database. The dataset includes 303 newly chartered commercial banks that opened their doors during 1985 and 2,350 established commercial banks that had been in operation for at least 25 years in 1985. The established banks each had less than $25 million in assets (1985 dollars); had equity capital equal to at least 5 percent of their assets; and were located in states in which at least four de novo banks started up in 1985. The dataset tracks each of these 2,653 banks across time and records the quarters in which banks left the dataset because they either failed or were acquired by another bank.

These data cover a period during which there were economic disruptions of sufficient magnitude to cause a statistically meaningful number of bank failures. Commercial bank failures were extremely rare in the U.S. during the 1950s,1960s, and 1970s, due to generally good economic times, regulatory limits on the risks that banks could take, and legal entry barriers that protected banks from competition. But the combination of banking deregulation and volatile interest rates during the 1970s and 1980s exposed banks to greater risks and more competition. As seen in figure 1, panel A, bank failures accelerated from near zero in 1980 to over 100 failures per year from the mid- 1980s through the early 1990s. The catalyst for these bank failures was a series of substantial economic disruptions, including a general recession in the early 1990s and a number of regional recessions in the mid-to late 1980s, the most disruptive of which was due to land price deflations in Texas and other oil-producing states.

For the purposes of this article, I consider a bank to have failed when at least one of the following occurs: 1) the bank is declared insolvent by its regulator; 2) the bank receives regulatory assistance (for example, a capital injection) without which it would become insolvent; or 3) the bank is acquired soon after its net worth has declined to less than 1 percent of assets. In terms of raw percentages, 16.5 percent of the de novo banks failed before the end of the 14year sample period. While this is over twice the 7.9 percent failure rate for the established banks in the sample, it is well below the reported failure rates for new (nonbank) business ventures. (See box 2 for a short discussion of new bank failures versus new business failures.)

(Table Omitted)

Captioned as: TABLE 2

Both the sample de novo banks and the sample established banks were more likely to be acquired than to fail during the sample period. The new banks were more likely to hold state charters; to be located in urban areas; to be located in the Southwest (primarily Texas, but also Louisiana and Oklahoma); and to be affiliates in multibank holding companies. Some of the hazard functions I estimate below include tests of whether these locational, organizational, and regulatory characteristics affect the probability of bank failure.

Nonparametric hazard functions

I use the bank failure data, summarized in table 2, to estimate separate hazard functions for newly chartered banks and established banks, and then compare these estimated hazard functions with the hypothetical hazard functions in figure 3. I employ two different hazard function techniques to produce these estimates-a nonparametric, or actuarial, approach, and a parametric, or duration model, approach.

An actuarial hazard function is simply a series of actuarial hazard rates strung together in chronological order. Calculating the actuarial hazard rates is straightforward and intuitive. For example, to calculate the 1990 hazard rate for a set of banks that were chartered in 1985, one simply divides the number of these banks that failed during 1990 by the number of the banks that still existed at the beginning of 1990. Thus, the hazard rate tells us the probability of failure in 1990 conditional on having survived for five years. The following, more exact, formula can be

used to calculate the actuarial hazard (Formula Omitted)

where n(t = 0) is the number of banks present at the beginning of the **analysis**; Ait) represents the number of these banks that failed during time period t; m(t) represents the number of these banks that were acquired in mergers during time period t, and T indicates the current time period. Note the subtle adjustment to the denominator in the second line of this formula: The denominator is reduced by one-half the number of banks that were acquired during the current time period. These banks clearly did not survive until the end of time period T, and subtracting some portion of these banks from the denominator acknowledges the possibility that they might have failed during time T had they not been acquired. Although weighting these banks by one-half is a crude and ad hoc adjustment, it is important to make some kind of adjustment because, as shown in table 2, acquired banks greatly outnumbered failed banks between 1985 and 1998.

I use the above formula to calculate 14 separate hazard rates (one rate for each of the 14 years from 1985 through 1998) for the 303 newly chartered banks. I repeat this exercise for the 2,350 established banks. Plotting the resulting hazard rates in chronological order generates two nonparametric hazard functions, which are displayed in figure 4.

In general, the nonparametric hazard functions in figure 4 resemble the hypothetical hazard functions posited in figure 3. The hazard rate for newly chartered banks is initially zero, and it remains below the established bank hazard rate for several years. As discussed above, this is most likely because the typical new bank holds a healthy equity cushion at the outset. After year three, the new bank hazard rate exceeds the established bank hazard rate, and it remains substantially higher than the established bank hazard rate until year eight. The hazard rate for newly chartered banks peaks in years five, six, and seven at about 1.2 percent-that is, if a newly chartered bank reaches the beginning of any of these years without failing or being acquired, it has about a 1.2 percent chance of failing before the year is out. At this point, the typical new bank's capital ratio has declined to established bank levels, but its profitability has not yet attained the level or degree of stability found at established banks. After year eight, the new bank hazard rate approaches the established bank hazard rate from above, suggesting that the maturation of new banks is well under way at this point.

The results displayed in figure 4 are consistent with the simple life-cycle theory of de novo bank failure. The nonparametric techniques used to generate figure 4 paint a good general picture of the rate at which new banks fail relative to established banks. But these nonparametric techniques do not control for the survivor bias in the data and, as a result, they understate the hazard rate at any given point in time. Furthermore, these techniques are not useful for testing how much, if any, of the difference between the new bank and established bank failure rates is caused by the economic, regulatory, and organizational conditions under which newly chartered banks operate. In the final step in this **analysis**, I use econometric duration **analysis** to estimate hazard functions. These parametric methods account for survivor bias and control for environmental conditions that can affect the probability of failure.

Parametric hazard functions

Duration **analysis** is a statistical regression approach. The dependent variable in these regressions is t, the length of time that passes between a new bank's start-up date and its subsequent failure. For established banks, t is the length of time between its first observation in the dataset (in this case, the first quarter of 1985) and its subsequent failure. The period measured by t is often referred to as a bank's duration. Because the banks in this dataset are observed quarterly, duration will range from t = for banks that fail during the quarter in which they begin operations, to t = 56 for banks that fail in the fourth quarter of 1998.

The simplest duration approach includes no explanatory variables. The

analyst starts by selecting a probability distribution formula that has a shape that is roughly similar to the actual distribution of the duration variable t, and uses maximum likelihood techniques to estimate parameter values that shape that probability distribution formula more exactly to the actual duration data. Here, I use a log-logistic distribution formula, because this is capable of producing hazard functions that have shapes similar to the hazard functions in figures 3 and 4. (Details of these duration model procedures can be found in the appendix to this article or in Greene, 1997). Once the parameters of the distribution formula have been estimated, they can be used to construct hazard functions as follows:
(Formula Omitted)

where f(T) is the probability that a bank fails at time T (that is, the log-logistic probability density) and F(T) is the probability that a bank fails before time T (that is, the log-logistic cumulative probability distribution). The denominator, 1 - F(T), is the log-logistic survival function, which is the probability that a bank neither fails nor is acquired before time T. This parametric hazard function has the same general interpretation as the nonparametric hazard function calculated in the previous section-they are both estimates of the probability that a bank will fail at time T given that it has survived until time T. One difference is that the hazard function generated by this parametric approach will be a smooth and continuous function of time similar to the hypothetical hazard functions in figure 3, as opposed to the segmented nonparametric hazard function in figure 4.

(Graph Omitted)

Captioned as: FIGURE 4

The duration models I estimate here control for survivor problems in the data. Recall that many of the sample banks either survived beyond the end of the sample period or were acquired during the sample period. These banks are known as censored observations. We cannot assign a duration **value** t to these banks because we cannot observe their ultimate fate (failure or survival). Furthermore, history suggests that very few of these banks will eventually fail, so including them in hazard rate calculations creates a downward bias by inflating the survival function 1 - F(t). Duration models can adjust for this problem by estimating the probability that censored banks will eventually fail, and then weighting the censored observations by this probability before estimating the parameters of the hazard function. (See the appendix for more details.)

The sample banks differ in terms of their geographic location, their organizational form, and their primary regulator. These characteristics could make a bank more or less likely to fail, or given that a bank does fail, these characteristics could influence how quickly it fails. For example, banks located in depressed economic regions will be more likely to fail and, absent regulatory intervention, will fail more quickly than banks located in economically healthy markets. Duration models can include a vector of independent variables, typically known as covariates, measuring the characteristics that vary across banks but remain constant for each bank over the sample period. I use a split population approach which estimates two regression coefficients for each of the covariates in the duration model. The first coefficient measures the covariate's impact on the probability that a bank will survive-a negative coefficient indicates that the covariate is associated with a lower probability of survival (higher probability of failure). The second coefficient measures the covariate's impact on a bank's duration-given that a bank will eventually fail, a negative coefficient indicates that the covariate is associated with a shorter duration (a faster failure).

The duration models I estimate include four covariates, each of which is expressed as a (0,) dummy variable. OCC = if the bank holds a federal charter (as opposed to a state charter). The OCC has traditionally practiced a more lenient chartering policy than most state chartering authorities, relying on market forces rather than administrative rules to determine the number of banks a market could support.7 A negative

coefficient on OCC would suggest that this policy caused new national banks to fail more often and/ or more quickly, on average, than new state-chartered banks. INDEPENDENT= 1 if the bank is either a freestanding business or a one-bank holding company (as opposed to being an affiliate of a multibank holding company) throughout the sample period. A negative coefficient on INDEPENDENT would suggest that banks not having access to the **financial** strength and managerial expertise of a multibank holding company tend to fail more often and/or more quickly. MSA = if the bank is located in an urban area. Banks in urban areas face greater competition than rural banks, but also may have greater opportunities for diversification. A negative coefficient on MSA would suggest that, on balance, these conditions cause banks in urban areas to fail more often and/or more quickly than rural banks. SW= 1 if the bank is located in the southwestern states of Louisiana, Texas, or Oklahoma, which experienced large numbers of bank failures during the mid- to late 1980s due to disruptions in energy-related industries. One would expect the coefficients on SW to be negative, reflecting lower survival probabilities and shorter duration times for banks in this region.

I add these four covariates to the duration model merely to illustrate how conditions and events external to the bank can affect its probability of failure and its time to failure. These four variables are not meant to be an exhaustive list of such conditions. Similarly, the duration model I estimate here is by no means definitive of the duration model techniques available to researchers. Other duration approaches do exist, including those that allow for time-varying covariates (for example, changes in economic, regulatory, or competitive conditions during each bank's duration). However, the multiple approaches I employ (including the Z-score and actuarial hazard function **analysis** conducted above) serve the purpose of this study, which is to test the simple life-cycle theory of de novo bank failure summarized in figure 3.

Table 3 displays the results of the duration models estimated separately for newly chartered banks and established banks. The estimated probability that the average bank will eventually fail is 19.65 percent for de novo banks and 8.93 percent for established banks. Note that these estimated failure probabilities are somewhat higher than the raw failure percentages shown at the bottom of table 2. In each case, the estimated probability is higher than the raw percentage because of the possibility that some of the censored observations will eventually fail.

Although established banks are less likely to fail, those that do fail have relatively short durations. Of the established banks that are expected to eventually fail, half of them will fail within an estimated 9.8 quarters (about 2.5 years) after the beginning of the sample period. Consistent with the life-cycle theory, newly chartered banks fail more slowly than established banks. It takes an estimated 21.1 quarters (about 5.25 years) for half of the de novo banks that are expected to fail to do so.

These differences in average duration can be seen clearly in figure 5, which charts the estimated hazard rates from the de novo and established bank duration models. Each of these functions is plotted based on the estimated coefficients shown in table 3 and the average values of the covariates for each sample. In general, these two estimated hazard functions resemble the shapes displayed above in figures 3 and 4. Thus, after controlling for censored data and a variety of environmental conditions, the failure patterns of newly chartered banks still differ substantially from the failure patterns of established banks.

The estimated probability of failure for established banks starts out above zero; peaks at about 8 percent for banks that survive for two years; and then slowly declines as the bank failure wave dissipates (see figure 1). In contrast, the estimated probability of failure for de novo banks starts out at zero and remains lower than the established bank hazard rate for three years; increases rapidly and peaks at nearly 14 percent for banks the survive for seven years; and then declines relatively quickly and begins to approach the established bank hazard rate. Note that both of

these hazard functions peak much higher on the vertical scale than did the actuarial hazard functions plotted in figure 4. Thus, by not controlling for censored observations and the overall low probability of eventual failure, the actuarial model substantially understated the hazard rates. Also, note that the hazard rates in figure 5 are in decline but are still positive at year 14, which reflects the non-zero probability of failure for the censored observations.

As expected, being located in one of the southwestern states reduces the probability of survival (or increases the probability of failure) for both de novo and established banks. Failing de novo banks also failed more quickly in this region, but failing established banks had longer than average durations. The latter result may indicate that regulators allowed troubled banks with longstanding business relationships (and, hence, more franchise **value**) more time to recover before stepping in to resolve them.8 Being located in a metropolitan statistical area reduced the probability of survival for de novo banks, but increased both the probability of survival for established banks and the survival time for established banks likely to eventually fail. Recall that intense competitive rivalry can cause banks to fail in urban markets, and that the lack of diversification opportunities can cause banks to fail in rural markets. The results suggest that these two phenomena affect de novo banks and established banks differently--on balance, de novo banks may be more sensitive to competition than to diversification risk, while small established banks may be more affected by a lack of diversification than by competitive rivalry. Being an independent bank or banking organization also reduces the probability of survival for de novo banks, which suggests that having access to the resources of multibank holding companies helps new banks survive. (I excluded this covariate from the established bank model because its presence prevented the model from converging.) The identity of a bank's primary regulator (OCC or state) is not a significant determinant of the probability of survival or the survival time for either set of banks.

Conclusion

Like all new business ventures, banks start with a business plan but no guarantee of success. So, despite the regulatory safeguards of on-site examinations, capital requirements, and other risk controls, we should not be surprised to find that new banks are more likely to fail than established banks. This article offers a simple framework that explains not only why but also when new banks are likely to fail.

(Table Omitted)

Captioned as: TABLE 3

(Graph Omitted)

Captioned as: FIGURE 5

My results suggest that the primary determinant of new bank failure is how new the bank is. Ironically, de novo banks are relatively unlikely to fail during their first few years of operation when they are earning negative profits. They are relatively more likely to fail during the years of positive profits that follow. Brand new, but unprofitable, banks are typically protected from failure by large initial capital cushions. However, equity cushions at de novo banks typically decline to established bank levels several years before their earnings become stable enough to justify these relatively low levels of capital.

What are the implications of this result for capital regulation at newly chartered banks? If ensuring a high rate of survival for de novo banks is a regulatory objective, then this result offers support for requiring high levels of start-up capital for new banks, and for holding young banks to higher capital requirements. Higher levels of required capital will make newly chartered banks less vulnerable to failure. Under such policies, de novo entrants might be a more credible long-run deterrent to market power in consolidating local markets. Indeed, in the wake of the wave of de novo bank failures during the 1980s, federal bank supervisory agencies did impose higher capital requirements on newly chartered banks.

On the other hand, promoting the safety and soundness of the banking system does not require that regulators prevent all bank failures, much less all failures of new banks. At some point, attempting to improve the survival rate of de novo banks by increasing the amount of capital necessary for investors to secure the charter will act as an entry barrier. Similarly, increasing the required capital ratios for young banks with charters already in hand will, at some point, depress investors' expected rates of return and discourage investment in new banks. Higher capital requirements for young banks could also slow the rate at which they can grow their balance sheets, hampering the beneficial impact of new banks in markets where existing banks (perhaps with market power) are not adequately serving the banking public.

What are the implications of this study for the bank chartering decision? During the period covered by this study, some state chartering authorities would approve or deny a charter application only after considering whether a local market "needed" an additional bank, based on the number of banks already serving the market and the expected rate of local economic growth. These restrictive chartering policies sought to reduce bank failure rates, and the **financial** disruptions that accompany them, by limiting competition in local banking markets. In contrast, the federal chartering authority practiced a liberal entry policy that explicitly ignored these "convenience and needs" issues, stressing instead the potential procompetitive benefits of de novo entry. My results indicate that the de novo national banks chartered in 1985 were no more likely to fail, or to fail quickly, than the de novo state banks chartered in that same year. This suggests that the benefits of a liberal chartering policy can be achieved without substantial increases in de novo bank failure rates. Additional research might confirm whether these findings, which are based on data from just 303 new banks chartered in a single year, also hold for banks chartered in other years and/or under different economic and regulatory circumstances.

(Table Omitted)

Captioned as: APPENDIX

(Table Omitted)

Sidebar:

BOX 1

Both the de novo bank sample and the established bank sample were taken from a primary data set used originally in a study by DeYoung and Hasan (1998). For the current study, I added variables from the "Reports of income and condition" (call reports). The primary dataset is an unbalanced panel consisting of 16,282 observations of 5,435 small, urban commercial banks at year-end 1988, 1990,1992, and 1994. Not all of the banks are present in each of the four years because some banks failed, were acquired, or received their charters during the sample period. There are 2,611 banks present in all four years, 977 banks in three of the four years, 1,005 banks in two years only, and 842 banks in just one ear.

Banks had to meet a number of conditions to be included in the primary dataset. First, banks had to have less than $500 million of assets (in 1994 dollars). By definition, newly chartered banks are small, and established banks that are large will not serve as good benchmarks against which to judge the progress of young banks. Large banks have access to production methods, risk strategies, distribution channels, and managerial talent not available to small banks. Second, all banks had to be headquartered in metropolitan statistical areas (MSAs). Demand for banking products, as well as competitive rivalry among banks, can be quite different in rural and urban markets, and may cause young banks to develop differently in these two environments. Third, banks had to be at least 12 months old at the time of observation. For example, a bank that was chartered during 1993, but was observed at year-end 1994, is referred to as a oneyear old bank. Brislin and Santomero (1991) find that **financial** statements are quite volatile during the first year of a bank's operations, which makes performance difficult to measure. Fourth, all banks had to make loans and

take deposits, eliminating special purpose banks such as credit card banks. Fifth, banks that were 14 years old or less (that is, banks that would be in the de novo sample) were excluded if they held more than $50 million in assets at the end of their first year. This filter prevents established banks that received new charters as part of regulatory reorganizations and established thrift institutions converting to bank charters from being identified as de novo banks.

The resulting de novo sample comprises 4,305 observations of 1,579 different banks 14 years old or younger. Roughly 47 percent of these de novo banks hold federal charters and roughly 21 percent are affiliates in multibank holding companies. The established bank sample was constructed by choosing 4,305 observations of 1,514 different banks, each more than 14 years old, from the primary dataset. Roughly 25 percent of these established banks hold federal charters and roughly 27 percent are affiliates in multibank holding companies. The established banks were chosen to have roughly the same asset-size distribution as the de novo bank sample, as follows: Banks more than 14 years old were grouped into ten asset categories ($0-$50 million, $50-$100 million, ..., $450-$500 million). Established banks were drawn at **random** from each of these size categories, depending on the number of de novo banks of each asset size. The assets of the resulting established bank sample average $55.97 million with a standard deviation of $49.64 million, compared with the de novo bank sample average of $54.39 million and standard deviation of $48.70 million.

Obviously, there is no bright line that separates de novo banks from established banks. I chose the 14-year old threshold for two reasons. First, it is the maximum age at which previous studies refer to commercial banks as de novo (see Huyser, 1986, and DeYoung and Hasan, 1998). Second, choosing a relatively large number for this threshold ensures that the maturity benchmarks in figure 2 contain only banks that are fully mature.

Sidebar:

BOX 2

During the 14 years covered by the bank failure dataset (see table 2), 16.5 percent of the newly chartered banks failed, compared with only 7.9 percent of the established banks of comparable size and location. To put these new bank failure rates into perspective, note that a 16.5 percent failure rate over 14 years is substantially lower than the failure rates typically reported for business start-ups in general. Raw data reported by the U.S. Small Business Administration (1992) suggest that at least 60 percent of new business ventures with less than 500 employees that started in 1977-78 failed within six years. Kirchhoff(1994, pp. 153-169) argues convincingly that these raw data overstate the new business failure rate because, among other things, the data in many instances define firms that changed owners or voluntarily shut down as having failed. After adjusting for these and other factors, Kirchhoff concludes that, in a best case scenario, 18 percent of new business ventures fail within eight years of start-up-about the same rate of failure as the de novo banks but in half the number of years. Furthermore, the 16.5 percent failure rate for new banks occurred during the worst period of bank failures since the Great Depression.

It should not be surprising that new banks have a better rate of survival than other new businesses. Both federal and state bank regulators deny charters to applicants with questionable **financial** credentials, restrict business activities, require high amounts of capital, apply regular scrutiny via on-site exams, and have the power to revoke bank charters. Banking start-ups face more severe entry barriers and ongoing scrutiny than new businesses in most other industries, and this selection bias naturally leads to a higher survival for new banks.

Reference:

REFERENCES

Reference:

Amel, Dean F., and J. Nellie Liang, 1997, "Determinants of entry and

profits in local banking markets," Review of Industrial Organization, Vol. 12, No. 1, pp 59-78.

Reference:

Berger, Allen N., Seth D. Bonime, Lawrence G. Goldberg, and Lawrence J. White, 1999, "The dynamics of market entry: The effects of mergers and acquisitions on de novo entry and customer service in banking," Proceedings from a Conference on Bank Structure and Regulation, Federal Reserve Bank of Chicago, forthcoming.

Reference:

Brewer III, Elijah, 1989, "Relationships between bank holding company risk and nonbank activity," Journal of Economics and Business, Vol. 41, November, pp. 337-353.

Brislin, Patricia, and Anthony Santomero, 1991, "De novo banking in the Third District," Business Review, Federal Reserve Bank of Philadelphia, January, pp. 3-12.

Reference:

Cole, Rebel A., and Jeffery W. Gunther, 1995, "Separating the likelihood and timing of bank failure," Journal of Banking and **Finance**, Vol. 19, No. 6, September, pp. 1073-1089.

DeYoung, Robert, Lawrence G Goldberg, and Lawrence J. White, 1999, "Youth, adolescence, and maturity of banks: Credit availability to small business in an era of banking consolidation," Journal of Banking and **Finance**, Vol. 23, No. 2-4, February, pp.463492.

Reference:

DeYoung, Robert, and Iftekhar Hasan, 1998, "The performance of de novo commercial banks: A profit efficiency approach," Journal of Banking and **Finance**, Vol. 22, May, pp. 565-587.

Greene, William H., 1997, Econometric **Analysis**, Upper Saddle River, NJ: Prentice Hall.

Reference:

Helwege, Jean, 1996, "Determinants of savings and loan failure," Journal of **Financial** Services Research, Vol.10, No. 4, pp. 373-392.

Hunter, William C. and Aruna Srinivasan, 1990, "Determinants of de novo bank performance," Economic Review, Federal Reserve Bank of Atlanta, March, pp. 14-25.

Hunter, William C., James A. Verbrugge, and David A. Whidbee, 1996, "Risk taking and failure in de novo savings and loans in the 1980s," Journal of **Financial** Services Research, Vol. lO, No. 3, pp. 235-272.

Reference:

Huyser, Daniel, 1986, "De novo bank performance in the Seventh District states," Banking Studies, Federal Reserve Bank of Kansas City, pp. 13-22.

Kirchhoff, Bruce A., 1994, Entrepreneurship and Dynamic Capitalism, Westport, CT and London: Greenwood, Praeger.

Reference:

Moore, Robert IL, and Edward C. Skelton, 1998, "New banks: Why enter when others exit?," **Financial** Industry Issues, Federal Reserve Bank of Dallas, First Quarter, pp. 1-7.

Seelig, Stephen, and Timothy Critchfield, 1999, "Determinants of de novo entry in banking," Federal Deposit Insurance Corporation, working paper, No. 99-1, January.

Reference:

U.S. Small Business Administration, 1992, The State of Small Business, Washington, DC: U.S. Government Printing Office.

Whalen, Gary, 1991, "A proportional hazards model of bank failure: An examination of its usefulness as an early warning tool," Economic Review, Federal Reserve Bank of Cleveland, Vol. 27, No. 1, First Quarter, pp. 2031.

Wheelock, David C., and Paul W. Wilson, 1995, "Explaining bank failures: Deposit insurance, regulation, and efficiency," Review of Economics and Statistics, November, pp. 689-700.

Author Affiliation:
Robert De Young is a senior economist and economic advisor at the Federal Reserve Bank of Chicago. The author thanks Iftekhar Hasan and Curt Hunter for their advice; Richard Cahill, Philip Jackson, and participants at a Federal Reserve Bank of Chicago seminar for helpful comments; Eli Brewer and David Marshall for reading an earlier draft of this article; and especially, Nancy Andrews for outstanding data support.

Special Features: Charts; Graphs; Equations; References
Classification: 8110 (CN=Commercial banking); 9190 (CN=United States); 4310 (CN=Regulation)
Descriptors: Commercial banks; Bank failures; Regulation of **financial** institutions
Geographic Names: US

36/9/4 (Item 4 from file: 268)
00363383
Bankrupt bank loan recoveries
Carty, Lea V; Hamilton, David T; Moss, Adam
Journal of Lending & Credit Risk Management , v 81 , n 10 , p 20-26 , Jun 1999 **Document Type:** Journal Article **ISSN:** 1088-7261 **Journal Code:** CBL **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 03560

Abstract:

A study by Moody's Investors Service, of interest to anyone holding bank loans in their portfolios, documents the bankruptcy and recovery experience of 200 bank loans. It extends and updates previous Moody's research by considering the nature, timing, and **value** of actual payouts to a sample of 200 bankrupt bank loan claims. The results of this study revealed heretofore unresearched aspects of the actual recoveries of the actual recoveries of bankrupt bank loans. Average recovery rate differences vis-a-vis previous research reflect the use of more detailed data and differing methodologies and samples. The average recovery rates for senior secured and unsecured bank loans estimated here are higher - 87% and 79% respectively - than recovery estimates previously for loans or bonds of similar seniority. These recovery rates were found to vary with both type of bankruptcy, and with the type and presence of collateral. The high recovery rates reflect the risk-mitigating features of many bank credit facilities - collateral, seniority and structure.

Text:

A study by Moody's Investors Service, of interest to anyone holding bank loans in their portfolios, documents the bankruptcy and recovery experience of 200 bank loans. It extends and updates previous Moody's research by considering the nature timing, and **value** of actual payouts to a sample of 200 bankrupt bank loan claims.

Increasingly, lenders desire to more accurately measure and manage credit risk in their loan portfolios; hence, their new interest in research quantifying that risk. Two principal components of credit risk are the probability of borrower default and the recovery. The equivalent of these are severity of loss and the economic **value** that lenders may expect for the defaulted obligation. While the likelihood of a default is roughly comparable for various debt obligations of the same obligor, these obligations are readily differentiated by the expected severity of the loss in the event a default occurs. The collateral and seniority of the obligation are two important determinants of the post-default recovery that bank loan investors may realize. These determinants ensure that the **value** of many defaulted bank loans is significantly different from that of other debt obligations.

This article characterizes the values, types, and dispersion of actual payouts to bankrupt bank credit facilities upon resolution of bankruptcy proceedings. Several summary measures of bankruptcy losses are distilled from these data, such as the average recovery, the dispersion or "risk" of recovery rates, and the average recovery rates by major collateral type. These risk measures should be of interest to lenders and risk professionals managing loan portfolios, pricing loans, analyzing structured instruments such as collateralized loan obligations, and in other riskrelated activities.

The **analysis** presented in this report is based on a unique dataset compiled by Moody's. In contrast to previous related research (which has principally relied on accounting or secondary market pricing data), the dataset underlying this research is based on a merger of bankruptcy resolution plans and Moody's database of bond and loan defaults. The detail available in the resolution plans allows each credit facility to be tracked through bankruptcy and its transformation into a postbankruptcy package of assets to be documented. The post-bankruptcy package of assets typically comprises some combination of debt, equity, cash, and other assets that can be valued and related to the original facility's claim. These data allow various aspects of the bankruptcy process to be examined, including the precise nature, **value**, and timing of actual payouts to bank credit facility claimants.

Methodology

Past studiesl of defaulted loan, bond, and preferred stock recoveries have considered market valuations for these instruments as recovery proxies. The recovery rates calculated in those reports are the ratios of the postdefault instrument values to their par amounts.

Using market valuations in this way has several benefits. It captures realized recoveries in cases where debtholders sold immediately after the default. It permits market participants to estimate recovery rates particular to their own circumstances. For example, Moody's has published recovery rates as the ratio of an instrument's post-default market **value** to its par **value**. This calculation would likely underestimate recoveries likely to be realized for investors purchasing distressed debt at a discount. The recovery rate these investors need might then be calculated-based on the price-based recovery data-as the ratio of the defaulted instrument's expected market price and its purchase price. The use of market pricing to estimate recoveries has the advantage of being practically implementable.

However, relying on market valuations as recovery indicators does not allow the timing, **value**, and nature of debtholders' actual recoveries to be described. Furthermore, any confidence in these prices as indicators of the eventual recovery is based on the efficiency of the market in which they are traded. Several researchers2 have found general, although sometimes qualified, support for the hypothesis that the bond market prices defaulted bonds efficiently There appears to be no similar research on the market for bankrupt loans, and the question of its efficiency is still unexplored. This research complements prior research on defaulted bank loan recoveries by examining in detail the actual payouts to

a sample of bankrupt loans.

The difficulties associated with this approach are significant.

Bankruptcy resolutions often satisfy creditor claims with a variety of debt, equity, and derivative instruments, many of which have no active secondary market in which the payments' true values can be determined. Because the payment amount, type, and maturity of bankruptcy claims is uncertain, accurate modeling of the instruments' **value** also is difficult. Additionally, many bankruptcies require a significant amount of time to resolve. Consequently, the time **value** of money can enter as a significant determinant of economic recoveries. However, these difficulties do not neutralize the **value** of research, and the need persists for a more complete understanding of the bankruptcy resolution process and its effects on corporate debt instruments.

This report concerns the estimation of the rate of recovery actually experienced by a sample of credit agreements. It analyzes the present **value** (as of the date of bankruptcy) of actual payouts associated with the bankruptcy's resolution. The recovery rate is defined as the present **value** of payouts presented as a fraction of the principal and interest due as of the date of filing. The three primary components required to calculate these recovery rates are:

1. Prepetition par **value** and accrued interest.
2. The **value**, nature and timing of any distribution returned to the creditor.

The discount rate to apply to each distribution.

Prepetition par **value** and accrued interest. Sources used to research the first component include:

Reorganization plan summaries

Moody's proprietary documents taken from its library of **financial** reports.

Press releases and press clippings.

Internal memoranda.

Moody's proprietary database of defaults and bankruptcies. Securities and Exchange Commission documents.

Value, nature and timing of distributions. The second component can be obtained only more subjectively because actual prices for many resolution payments are not available. Payments awarded to each creditor class varied widely across reorganizations and included such **financial** instruments as cash, debt, equity, derivatives, and enhancements to the terms of any surviving liability. It was not possible to document the nature and timing of a majority of these payments, but valuing them required additional subjective **analysis**. A number of approaches were used to **value** these items as consistently as possible across each bankruptcy resolution in our sample.

The preferred approach to valuing resolution distributions was to rely on market pricing after the bankruptcy's resolution. Primary sources for obtaining equity, preferred, warrant, and debt prices were Bloomberg, IDC, Citibank, Goldman Sachs, BDS Securities, Lehman Brothers, Merrill Lynch, Loan Pricing Corporation, the borrower's **financial** statements, and the libraries of certain domestic stock exchanges. When possible, prices taken were as of the borrower's exit date from bankruptcy However, when a quote was not available immediately after the borrower had exited bankruptcy, the price of the new security was discounted by the total return of an applicable market benchmark.

In cases where reliable pricing was not available, a discounted cash flow approach was used to **value** certain distributions. Because of its limitations, this approach was used only to **value** new or amended debt instruments. A **historical** Moody's rating or an estimated rating (based on a Moody's analyst's opinion of the reorganized borrower's credit quality) on the new or amended security was used to derive the discount rate. The discount rate was the median yield for similarly rated debt instruments in the market at the time.

In certain cases, some payments were valued based on appraisals by

Moody's analysts and independent, qualified agents. While the subjectivity and limitations of this technique may have potential effects on the study, Moody's **historical** expertise in securities **analysis** and the bankruptcy process limits any distortions introduced and enables us to conduct a meaningful study of loan recoveries.

Discount rate. The third component, the discount rate, is defined in relation to the terms of the original credit agreement. Hence, the discount rate applied to any loan workout is the contractual lending rate. In many cases, it was possible to determine this rate precisely. In others, estimates drawn from analyses of the credit environment and similarly structured instruments were used.

Data

This report is based on a dataset of 200 bank loans involved with the bankruptcies of 119 large, public borrowers. Most of these borrowers filed for regular Chapter 11 protection, and a significant portion of those (29%) filed prepackaged Chapter lls. The earliest bankruptcy in this sample is LTV Corporation's July 17, 1986 Chapter 11 filing. The most recent is Payless Cashways, Inc.'s July 21, 1997 Chapter 11. The bankruptcies analyzed in this report cover the second half of the 1980s, the recession of the early 1990s (and the coincident junk bond market crash) as well as the boom years since. Figure 1 shows that bankruptcies were most frequent in the early 1990s.

Figure 1 also highlights the increased relative popularity of prepackaged Chapter lls since the early 1990s. Chapter 11 has the benefits of protecting debtors from creditors but is expensive. Informal workouts, while less costly, are marred by collective action problems. The 1990 Revenue Reconciliation Act made them even less attractive by making income realized from debt forgiveness taxable. Prepackaged bankruptcies take advantage of the Chapter 11 process to mitigate the holdout problem and of pre-filing negotiations to speed the process to completion, thus limiting costs.

(Graph Omitted)

Captioned as: Figure 1

Figure 2 details the industrial cross-section of the firms in the sample. Industrials, retailers and consumer products companies accounted for 58%. No other single industry accounted for more than 7% of the sample. This pattern is roughly consistent with Moody's database of bond defaults in the same period.

The Length of Time Spent in Bankruptcy

The length of time that a bankruptcy takes to resolve is a critical aspect of recovery. The longer the recovery, the greater the cost to unsecured debtholders whose claims to interest do not typically accrue during the course of a workout. Even secured debtholders are not always paid interest, although it may accrue during the course of the bankruptcy's resolution. Hence, even if the debtholders are confident that their claim will be satisfied, the uncertain timing of settlement may limit the appeal of bankrupt debt to many investors-especially those seeking current income. Consequently, investors require a compensatory liquidity premium.

Figure 3 describes the overall distribution of the lengths of time spent in bankruptcy and breaks them out by bankruptcy type. Overall, the shortest time to resolution of the bankruptcies in the sample is a little over one month for the prepackaged Chapter 11 of Memorex Telex Corporation. The maximum is almost 7 years for LTV Corporation and the average is about one year and 3 months. The distribution of time to recovery, however, varies in the expected way with the bankruptcy type. Prepackaged bankruptcies require less time on average to resolve, about 2.2 months, compared to 1.6 years on average for regular Chapter 11 filings-almost 8 times longer. Furthermore, the distribution of times to resolution varies more widely for Chapter 11 filings than for prepackaged Chapter ls. The longest time any of the borrowers that filed for a prepackaged Chapter 11 spent in bankruptcy is six months, compared to almost seven years for Chapter 11 filers. Furthermore, the dispersion of the distribution as

measured by the standard deviation is much greater for Chapter ll Is.

The average of the lengths of times to resolution may overstate the amount of time one can expect a typical filing to take to resolve. This is because the distribution of times to resolution is necessarily truncated at zero. The median is a more reliable measure of the center of similarly skewed distributions' centers. In each case shown in Figure 3, the median length of time to resolution is less than the average reflecting this fact.

The Nature of Typical Bankruptcy Resolutions

Almost by definition, a bankruptcy resolution requires the relaxation of the current **financial** commitments of the firm. The **financial** reorganization typically requires extending debt maturity, eliminating, diminishing, deferring, or otherwise modifying interest payable (for example, making pay-in-kind); and diminishing the principal or replacing the claim with contracts that place less strain on the firm's short-term contractual cash flow (for example, preferred stock, equity, warrants, right, options, or other derivative instruments).

(Graph Omitted)

Captioned as: Figure 2

(Graph Omitted)

Captioned as: Figure 3

The pie chart in Figure 4 summarizes how often the claims of bankrupt bank loans are satisfled with various types of **financial** instruments. To make the pie chart, we sorted by instrument type all of the recoveries earned on each of the 200 bank loans in this study by instrument type. For example, 43.8% of the recovery on all 200 loans took the form of bank loans. New loans typically have terms that reduce the strain on the current cash flow of the borrowing firm (such as extended maturities or reduced interest rates), but new loans may also enjoy an enhanced collateral position. Bank loans and cash together account for 80.2% of the total recovery analyzed in this report. Long-term public and private debt account for another 6.5%. "Other" also accounts for 6.5% and is composed of various types of assets, including collateral or income generated by legal settlements. Equity makes up another 6.3%, leaving preferred stock, rights, and warrants to make up the remaining 0.4%.

Bankrupt Bank Loan Recovery Rates

Figures 5 and 6 display the average estimated recovery rates for senior secured and senior unsecured bankrupt bank loans. Figure 7 provides more detail on the distribution of the recovery rates. The additional protection afforded by security shows up here as a higher average recovery-87%for 178 senior secured bank loans. For senior unsecured bank loans, the average recovery falls to 79%.

However, the average recovery is only one measure of what could be considered the most "typical **value**" that would be realized for a bankrupt bank loan. The distributions of recoveries are skewed towards the high end of the scale. This indicates that while the average recovery of senior secured bank loans is 87%, more than half are greater. The median recovery rate, 100%, is a more robust measure of the center of an asymmetric distribution. It indicates that a loan drawn at **random** from this sample would yield a recovery of at least 100% half of the time.

The tremendous dispersions of the recovery rate distributions are an important characteristic. Lenders to the supermarket Almac's, which filed for creditor protection in August 1993, recovered 7.4%, the lowest recovery calculated for senior claims. The greatest is well over 100%. One of the most common measures of dispersion is standard deviation, which is 23% for senior secured loans and 27% for senior unsecured loans. The size of these standard deviations means that a loan may well see a recovery rate much greater or much lower than the average.

(Graph Omitted)

Captioned as: Figure 4

(Graph Omitted)

Captioned as: Figure 5

While somewhat useful, confidence **intervals** derived from averages and standard deviations can be misleading when they're applied to small samples from skewed distributions. An alternative is to consider a band into which approximately 90% of the observations actually fall. In Moody's sample of 178 senior secured recovery rates, 160 (90%) fell between the tenth lowest recovery rate, 36%, and the tenth highest recovery rate, 100%. This distribution-free methodology shows that recovery rates of less than 36% and greater than 100% are relatively rare. While actual recoveries on senior secured bank debt are high, both on average and relative to public bonds, uncertainty over actual **value** is tremendous. Even though the median recovery **value** is 100% of principal and interest due on senior secured bank loans, recoveries as low as 36% are not uncommon.

Figure 7 breaks out recoveries more descriptively by Chapter 11 and prepackaged Chapter 11. For both the senior secured and senior unsecured bank loans, the average recovery rate is higher for prepackaged Chapter lls than for Chapter lls, and the median recovery rate is at least as large. Moreover, the sample of prepackaged Chapter lls for the senior secured bank loans is so big, it permits calculation of the sample standard deviation: 16%. This is lower than for senior secured Chapter lls, and it indicates that senior secured bank loan holders face less risk if a prepackaged Chapter 11 is filed. This pattern is apparent in part because the senior and secured position of these lenders protects them relatively more in the event of a Chapter 11 filing. If creditors were faced with a prepackaged Chapter 11 that did hurt senior secured bank lenders more than a regular Chapter 11 could be expected to, the bank creditors would opt for a regular Chapter 11 filing over the prepack. Hence, there is a bias towards prepacks that benefit senior secured bank lenders.

The **Value** of Collateral

An earlier report by Stumpp, Marshella, Mulvaney, and Hilderman (1997) stressed the belief that not all collateral gives equal security. Moody's new data support that conclusion. Figure 8 presents the results of determining and broadly classifying the supporting collateral for 136 of the bankrupt senior secured bank loans, which in turn permitted average recovery rates segmented by collateral type to be estimated.

The results are consistent with intuition, and the opinion of Stumpp et al that not all collateral gives equal security. The most liquid collateral-accounts receivable, cash and inventory-produces the highest average recovery rate. The stock of subsidiaries is the collateral associated with the lowest average recovery rate.

(Graph Omitted)

Captioned as: Figure 6

(Chart Omitted)

Captioned as: Figure 7

Stock pledges offer limited benefits because pledged stock is illiquid and often adversely affected by the bankruptcy filing. Furthermore, the lender's claim is secured only to the extent of the equity **value** of the entity that has pledged its stock. The subsidiaries of a parent filing for bankruptcy are likely to be insolvent or underperforming. It may surprise some readers that the stock of subsidiaries generate an average recovery rate as high as 74%. Keep in mind two factors when interpreting these statistics: 1. Most of this sample was formed during a bull market in equities, and the stock of subsidiaries may have been more valuable then than it would be under more typical equity market scenarios.

2. There is tremendous uncertainty surrounding the average **value** as measured by the sample standard deviation. Interestingly, the dispersion of the recovery rates - measured here by their standard deviation - tends to fall as average recovery rates rise. This pattern suggests that as the expected recovery rate of a loan falls, the loan's risk-the possibility for any particular loan's recovery to differ from the average by a wide margin-increases. Conversely, to the extent that all loans will behave like those in this sample, loans backed

by accounts receivable, cash or inventory will have a greater recovery rate than loans backed by the stock of subsidiaries, and recovery will be more certain.

Conclusion

The results of this study revealed heretofore unresearched aspects of the actual recoveries of bankrupt bank loans. Average recovery rate differences vis-a-vis previous research reflect our use of more detailed data and differing methodologies and samples. The average recovery rates for senior secured and unsecured bank loans estimated here are higher-87% and 79% respectively-than recovery rates estimated previously for loans or bonds of similar seniority.

These recovery rates were found to vary with both type of bankruptcy (Chapter 11 and prepackaged Chapter 11), and with the type and presence of collateral. The high recovery rates reflect the risk-mitigating features of many bank credit facilities-collateral, seniority and structure. 0

(Chart Omitted)

Captioned as: Figure 8

Footnote:

Footnotes

Footnote:

1 Carty (1994); Carty & Lieberman (1996); Keenan, Carty & Shtogrin (1998); Keenan, Carty, Shtogrin & Fons (1998).

2 Eberhart & Sweeny (1992); Altman & Eberhart (1994); Warner (1977).

Reference:

n.irrc

Reference:

Altman, Edward I. and Eberhart, Allan C. (1994), "Do Seniority Provisions Protect Bondholders' Investments?," A New York University Salomon Center Working Paper, Series S-94-12. Carty, Lea V. (1994), "Preferred Stock Dividend and Credit Risk," A Moody's Special Report, December 1994.

Reference:

Eberhart, Allan C. and Sweeney, Richard J. (1992), "Does the Bond Market Predict Bankruptcy Settlements?," The Journal of **Finance**, Vol. XLVII, No. 3, July 1992.

Keenan, Sean C., Carty, Lea V., Shtogrin, Igor (1998), "**Historical** Default Rates of Corporate Bond Issuers, 1920 - 1997," A Moody's Special Report, March, 1998. Also, yearly updates from 1994 through 1997.

Reference:

Keenan, Sean C., Carty, Lea V., Shtogrin, Igor, and Fons, Jerome S. (1998), "Preferred Stock Ratings and Credit Risk," A Moody's Special Report, March 1998.

Reference:

Stumpp, Pamela, Marshella, Tom, Mulvaney Michael, and Hilderman Mara (1997), "A Sense of Security: Moody's Approach to Evaluating Bank Loan Structure and Collateral," A Moody's Special Report, October 1997.

Warner, Jerold B., 1977, "Bankruptcy, Absolute Priority, and the Pricing of Risky Debt Claims," Journal of **Financial** Economics, 1977, Vol. 4, No. 3, 239-276.

Author Affiliation:

1999 by RMA and Moody s Investors Service, Inc. Cart3, is managing director ofMoody s Investors Senvice, Inc. Carty is managing director of Moody's Investors Services Risk Management group. Hamilton is an associate analyst in Moody's Investors Services Risk Management group. Moss is an associate analyst in Moody's Investors Services Speculative Grade group.

Author Affiliation:

This article was extracted from a longer and more detailed Moody's report by the same title. To obtain a copy of the study or to get more information, call Mfoody y's client services desk at 212-553-1653.

Special Features: Charts; Graphs; References
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 8100 (CN=Financial services industry); 3200 (CN=Credit management)
Descriptors: Studies; Bank loans; **Portfolio** management; Bankruptcy; Credit management
Geographic Names: US

36/9/5 (Item 5 from file: 268)
00348495
Industry practices in credit risk modeling and internal capital allocations: Implications for a Models-based regulatory capital standard
Jones, David; Mingo, John
Economic Policy Review , v 4 , n 3 , p 53-60 , Oct 1998 **Document Type:** Journal Article **Journal Code:** EPV **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 04186

Abstract:

Information from internal credit risk models might be usefully incorporated into regulatory or supervisory capital policies. Internal capital allocations against credit risk are based on a bank's estimate of the probability density function (PDF) for credit losses. Credit risk models are used to estimate these PDFs. Four aspects of credit risk modeling are: 1. the conceptual framework, 2. credit-related optionality, 3. model calibrations, and 4. model validation. Incorporating internal credit risk measurement and capital allocation systems into the supervisory and/or regulatory framework will occur neither quickly nor without significant difficulties. The current one-size-fits-all system of risk-based capital requirements increasingly is inadequate to the task of measuring large bank soundness.

Text:

I. WHY SHOULD REGULATORS BE INTERESTED IN CREDIT RISK MODELS?

Bank supervisors have long recognized two types of shortcomings in the Basle Accord's risk-based capital (RBC) framework. First, the regulatory measures of "capital" may not represent a bank's true capacity to absorb unexpected losses. Deficiencies in reported loan loss reserves, for example, could mask deteriorations in banks' economic net worth. Second, the denominator of the RBC ratios, total risk-weighted assets, may not be an accurate measure of total risk. The regulatory risk weights do not reflect certain risks, such as interest rate and operating risks. More importantly, they ignore critical differences in credit risk among **financial** instruments (for example, all commercial credits incur a 100 percent risk weight), as well as differences across banks in hedging, **portfolio** diversification, and the quality of risk management systems.

These anomalies have created opportunities for "regulatory capital arbitrage" that are rendering the formal RBC ratios increasingly less meaningful for the largest, most sophisticated banks. Through securitization and other **financial** innovations, many large banks have lowered their RBC requirements substantially without reducing materially their overall credit risk exposures. More recently, the September 1997 Market Risk Amendment to the Basle Accord has created additional arbitrage opportunities by affording certain credit risk positions much lower RBC requirements when held in the trading account rather than in the banking book.

Given the prevalence of regulatory capital arbitrage and the unstinting pace of **financial** innovation, the current Basle Accord may soon become overwhelmed. At least for the largest, most sophisticated banks, it seems clear that regulators need to begin developing the next generation of capital standards now-before the current framework is completely outmoded. "Internal models" approaches to prudential regulation are presently the only long-term solution on the horizon.

The basic problem is that securitization and other forms of capital arbitrage allow banks to achieve effective capital requirements well below the nominal 8 percent Basle standard. This may not be a concern-indeed, it may be desirable from a resource allocation perspective-when, in specific instances, the Basle standard is way too high in relation to a bank's true risks. But it is a concern when capital arbitrage lowers overall prudential standards. Unfortunately, with the present tools available to supervisors, it is often difficult to distinguish these cases, especially given the lack of transparency in many offbalance-sheet credit positions.

Ultimately, capital arbitrage stems from the disparities between true economic risks and the "one-sizefits-all" notion of risk embodied in the Accord. By contrast, over the past decade many of the largest banks have developed sophisticated methods for quantifying credit risks and internally allocating capital against those risks. At these institutions, credit risk models and internal capital allocations are used in a variety of management applications, such as risk-based pricing, the measurement of risk-adjusted profitability, and the setting of **portfolio** concentration limits.

II. THE RELATIONSHIP BETWEEN PDF AND ALLOCATED ECONOMIC CAPITAL

Before discussing various credit risk models per se, it may be helpful to describe how these models are used within banks' capital allocation systems. Internal capital allocations against credit risk are based on a bank's estimate of the probability density function (PDF) for credit losses. Credit risk models are used to estimate these PDFs (see chart). A risky **portfolio** is one whose PDF has a relatively long, fat tail-that is, where there is a significant likelihood that actual losses will be substantially higher than expected losses, shown as the left dotted line in the chart. In this chart, the probability of credit losses exceeding the level X is equal to the shaded area under the PDF to the right of X.

The estimated capital needed to support a bank's credit risk exposure is generally referred to as its "economic capital" for credit risk. The process for determining this amount is analogous to VaR methods used in allocating economic capital against market risks. Specifically, the economic capital for credit risk is determined in such a way that the estimated probability of unexpected credit losses exhausting economic capital is less than the bank's "target insolvency rate." Capital allocation systems generally assume that it is the role of reserving policies to cover expected credit losses, while it is the role of equity capital to cover credit risk, or the uncertainty of credit losses. Thus, required economic capital is the amount of equity over and above expected losses necessary to achieve the target insolvency rate. In the chart, for a target insolvency rate equal to the shaded area, the required economic capital equals the distance between the two dotted lines.

In practice, the target insolvency rate is usually chosen to be consistent with the bank's desired credit rating. For example, if the desired credit rating is AA, the target insolvency rate might equal the **historical** one-year default rate for AA-rated corporate bonds (about 3 basis points).

To recap, economic capital allocations for credit risk are based on two critical inputs: the bank's target insolvency rate and its estimated PDF for credit losses. Two banks with identical portfolios, therefore, could have very different economic capital allocations for credit risk, owing to differences in their attitudes toward risk taking, as reflected in their target insolvency rates, or owing to differences in their methods for estimating PDFs, as reflected in their credit risk models. Obviously, for competitive equity and other reasons, regulators prefer to apply the same minimum soundness standard to all banks. Thus, any internal models approach to regulatory capital would likely be based on a bank's estimated PDF, not on the bank's own internal economic capital allocations. That is, the regulator would likely (a) decide whether the bank's PDF estimation process was acceptable and (b) at least implicitly, set a regulatory maximum insolvency probability (rather than accept the bank's target insolvency rate if such a rate was deemed "too high" by regulatory standards).

(Graph Omitted)

Captioned as: THE RELATIONSHIP BETWEEN PDF AND ALLOCATED ECONOMIC CAPITAL LOSSES

(Chart Omitted)

Captioned as: OVERVIEW OF RISK MEASUREMENT SYSTEMS

III. TYPES OF CREDIT RISK MODELS

When estimating the PDF for credit losses, banks generally employ what we term either "top-down" or "bottom-up" methods (see exhibit). Top-down models are often used for estimating credit risk in consumer or small business portfolios. Typically, within a broad subportfolio, such as credit cards, all loans would be treated as more or less homogeneous. The bank would then base its estimated PDF on the **historical** credit loss rates for that subportfolio taken as a whole. For example, the variance in subportfolio loss rates over time could be taken as an estimate of the variance of loss rates associated with the current subportfolio. A limitation of top-down models, however, is that they may not be sensitive to changes in the subportfolio's composition. That is, if the quality of the bank's card customers were to change over time, PDF estimates based on that **portfolio**'s **historical** loss rates could be highly misleading.

Where changes in **portfolio** composition are a significant concern, banks appear to be evolving toward bottom-up models. This is already the predominant method for measuring the credit risks of large and middlemarket customers. A bottom-up model attempts to quantify credit risk at the level of each individual loan, based on an explicit credit evaluation of the underlying customer. This evaluation is usually summarized in terms of the loan's internal credit rating, which is treated as a proxy for the loan's probability of default. The bank would also estimate the loan's loss rate in the event of default, based on collateral and other factors. To measure credit risk for the **portfolio** as a whole, the risks of individual loans are aggregated, taking into account correlation effects. Unlike top-down methods, therefore, bottom-up models explicitly consider variations in credit quality and other compositional effects.

IV. MODELING ISSUES

The remainder of this summary focuses on four aspects of credit risk modeling: the conceptual framework, credit-related optionality, model calibrations, and model validation. The intent is to highlight some of the modeling issues that we believe are significant from a regulator's perspective; the full version of our paper provides significantly greater detail.

A. CONCEPTUAL FRAMEWORK

Credit risk modeling procedures are driven importantly by a bank's underlying definition of "credit losses" and the "planning horizon" over which such losses are measured. Banks generally employ a one-year planning horizon and what we refer to as either a default-mode (DM) paradigm or a mark-to-market (MTM) paradigm for defining credit losses.

1. Default-Mode Paradigm

At present, the default-mode paradigm is by far the most common approach to defining credit losses. It can be thought of as a representation of the traditional "buyand-hold" lending business of commercial banks. It is sometimes called a "two-state" model because only two outcomes are relevant: nondefault and default. If a loan does not default within the planning horizon, no credit loss is incurred; if the loan defaults, the credit loss equals the difference between the loan's book **value** and the present **value** of its net recoveries.

2. Mark-to-Market Paradigm

The mark-to-market paradigm generalizes this approach by recognizing that the economic **value** of a loan may decline even if the loan does not formally default, This paradigm is "multi-state" in that "default" is only one of several possible credit ratings to which a loan could migrate. In effect, the credit **portfolio** is assumed to be marked to market or, more accurately, "marked to model." The **value** of a term loan, for example, typically would employ a discounted cash flow methodology, where the credit spreads used in valuing the loan would depend on the instrument's credit rating.

To illustrate the differences between these two paradigms, consider a loan having an internal credit rating equivalent to BBB. Under both paradigms, the loan would incur a credit loss if it were to default during the planning horizon. Under the mark-to-market paradigm, however, credit losses could also arise if the loan were to suffer a downgrade short of default (such as migrating from BBB to BB) or if prevailing credit spreads were to widen. Conversely, the **value** of the loan could increase if its credit rating improved or if credit spreads narrowed.

Clearly, the planning horizon and loss paradigm are critical decision variables in the credit risk modeling process. As noted, the planning horizon is generally taken to be one year. It is often suggested that one year represents a reasonable **interval** over which a bank-in the normal course of business-could mitigate its credit exposures. Regulators, however, tend to frame the issue differently-in the context of a bank under stress attempting to unload the credit risk of a significant **portfolio** of deteriorating assets. Based on experience in the United States and elsewhere, more than one year is often needed to resolve asset-quality problems at troubled banks. Thus, for the banking book, regulators may be uncomfortable with the assumption that capital is needed to cover only one year of unexpected losses.

Since default-mode models ignore credit deteriorations short of default, their estimates of credit risk may be particularly sensitive to the choice of a one-year horizon. With respect to a three-year term loan, for example, the one-year horizon could mean that more than two-thirds of the credit risk is potentially ignored. Many banks attempt to reduce this bias by making a loan's estimated probability of default an increasing function of its maturity. In practice, however, these adjustments are often made in an ad hoc fashion, so it is difficult to assess their effectiveness.

B. CREDIT-RELATED OPTIONALITY

In contrast to simple loans, for many instruments a bank's credit exposure is not fixed in advance, but rather depends on future (**random**) events. One example of such "creditrelated optionality" is a line of credit, where optionality reflects the fact that drawdown rates tend to increase as a customer's credit quality deteriorates. As observed in connection with the recent turmoil in foreign exchange markets, credit-related optionality also arises in derivatives transactions, where

counterparty exposure changes **randomly** over the life of the contract, reflecting changes in the amount by which the bank is "in the money."

As with the treatment of optionality in VaR models, credit-related optionality is a complex topic, and methods for dealing with it are still evolving. At present, there is great diversity in practice, which frequently leads to very large differences across banks in credit risk estimates for similar instruments. With regard to virtually identical lines of credit, estimates of stand-alone credit risk can differ as much as a tenfold. In some cases, these differences reflect modeling assumptions that, quite frankly, seem difficult to justify-for example, with respect to committed lines of credit, some banks implicitly assume that future drawdown rates are independent of future changes in a customer's credit quality. Going forward, in our view the treatment of credit-related optionality needs to be a priority item, both for bank risk modelers and their supervisors.

C. MODEL CALIBRATION

Perhaps the most difficult aspect of credit risk modeling is the calibration of model parameters. To illustrate this process, note that in a default-mode model, the credit loss for an individual loan reflects the combined influence of two types of risk factors-those determining whether or not the loan defaults and, in the event of default, risk factors determining the loan's loss rate. Thus, implicitly or explicitly, the model builder must specify (a) the expected probability of default for each loan, (b) the probability distribution for each loan's loss-rate-given-default, and (c) among all loans in the **portfolio**, all possible pair-wise correlations among defaults and loss-rates-given-default. Under the mark-to-market paradigm, the estimation problem is even more complex, since the model builder needs to consider possible credit rating migrations short of default as well as potential changes in future credit spreads.

This is a daunting task. Reflecting the longer term nature of credit cycles, even in the best of circumstancesassuming parameter stability-many years of data, spanning multiple credit cycles, would be needed to estimate default probabilities, correlations, and other key parameters with good precision. At most banks, however, data on **historical** loan performance have been warehoused only since the implementation of their capital allocation systems, often within the last few years. Owing to such data limitations, the model specification process tends to involve many crucial simplifying assumptions as well as considerable judgment.

In our full paper, we discuss assumptions that are often invoked to make model calibration manageable. Examples include assumptions of parameter stability and various forms of independence within and among the various types of risk factors. Some specifications also impose normality or other parametric assumptions on the underlying probability distributions.

It is important to note that estimation of the extreme tail of the PDF is likely to be highly sensitive to these assumptions and to estimates of key parameters. Surprisingly, in practice there is generally little **analysis** supporting critical modeling assumptions. Nor is it standard practice to conduct sensitivity testing of a model's vulnerability to key parameters. Indeed, practitioners generally presume that all parameters are known with certainty, thus ignoring credit risk issues arising from parameter uncertainty or model instability. In the context of an internal models approach to regulatory capital for credit risk, sensitivity testing and the treatment of parameter uncertainty would likely be areas of keen supervisory interest.

D. MODEL VALIDATION

Given the difficulties associated with calibrating credit risk models, one's attention quickly focuses on the need for effective model validation procedures. However, the same data problems that make it difficult to calibrate these models also make it difficult to validate the models. Owing to insufficient data for out-of-sample testing, banks generally do not conduct statistical back testing on their estimated PDFs.

Instead, credit risk models tend to be validated indirectly, through various market-based "reality" checks. Peer-group **analysis** is used extensively to gauge the reasonableness of a bank's overall capital allocation process. Another market-based technique involves comparing actual credit spreads on corporate bonds or syndicated loans with the break-even spreads implied by the bank's internal pricing models. Clearly, an implicit assumption of these techniques is that prevailing market perceptions and prevailing credit spreads are always "about right."

In principle, stress testing could at least partially compensate for shortcomings in available back-testing methods. In the context of VaR models, for example, stress tests designed to simulate hypothetical shocks provide useful checks on the reasonableness of the required capital levels generated by these models. Presumably, stress-testing protocols also could be developed for credit risk models, although we are not yet aware of banks actively pursuing this approach.

V. POSSIBLE NEAR-TERM APPLICATIONS OF CREDIT RISK MODELS

While the reliability concerns raised above in connection with the current generation of credit risk models are substantial, they do not appear to be insurmountable. Credit risk models are progressing so rapidly it is conceivable they could become the foundation for a new approach to setting formal regulatory capital requirements within a reasonably near time frame. Regardless of how formal RBC standards evolve over time, within the short run supervisors need to improve their existing methods for assessing bank capital adequacy, which are rapidly becoming outmoded in the face of technological and **financial** innovation. Consistent with the notion of "risk-focused" supervision, such new efforts should take full advantage of banks' own internal risk management systems-which generally reflect the most accurate information about their credit exposuresand should focus on encouraging improvements to these systems over time.

Within the relatively near term, we believe that there are at least two broad areas in which the inputs or outputs of bank's internal credit risk models might usefully be incorporated into prudential capital policies. These include (a) the selective use of internal credit risk models in setting formal RBC requirements against certain credit positions that are not treated effectively within the current Basle Accord and (b) the use of internal credit ratings and other components of credit risk models for purposes of developing specific and practicable examination guidance for assessing the capital adequacy of large, complex banking organizations.

A. SELECTIVE USE IN FORMAL RBC REQUIREMENTS

Under the current RBC standards, certain credit risk positions are treated ineffectually or, in some cases, ignored altogether. The selective application of internal risk models in this area could fill an important void in the current RBC framework for those instruments that, by virtue of their being at the forefront of **financial** innovation, are the most difficult to address effectively through existing prudential techniques.

One particular application is suggested by the November 1997 Notice of Proposed Rulemaking on Recourse and Direct Credit Substitutes (NPR) put forth by the U.S. banking agencies. The NPR discusses numerous anomalies regarding the current RBC treatment of recourse and other credit enhancements supporting banks' securitization activities. In this area, the Basle Accord often produces dramatically divergent RBC requirements for essentially equivalent credit risks, depending on the specific contractual form through which the bank assumes those risks.

To address some of these inconsistencies, the NPR proposes setting RBC requirements for securitization-related credit enhancements on the basis of credit ratings for these positions obtained from one or more accredited rating agencies. One concern with this proposal is that it may be costly for banks to obtain formal credit ratings for credit enhancements that currently are not publicly rated. In addition, many large banks already produce internal credit ratings for such instruments, which, given the

quality of their internal control systems, may be at least as accurate as the ratings that would be produced by accredited rating agencies. A natural extension of the agencies' proposal would permit a bank to use its internal credit ratings (in lieu of having to obtain external ratings from accredited rating agencies), provided they were judged to be "reliable" by supervisors.

A further extension of the agency proposal might involve the direct use of internal credit risk models in setting formal RBC requirements for selected classes of securitization-related credit enhancements. Many current securitization structures were not contemplated when the Accord was drafted, and cannot be addressed effectively within the current RBC framework. Market acceptance of securitization programs, however, is based heavily on the ability of issuers to quantify (or place reasonable upper bounds on) the credit risks of the underlying pools of securitized assets. The application of internal credit risk models, if deemed "reliable" by supervisors, could provide the first practical means of assigning economically reasonable capital requirements against such instruments. The development of an internal models approach to RBC requirements-on a limited scale for selected instrumentsalso would provide a useful test bed for enhancing supervisors' understanding of and confidence in such models, and for considering possible expanded regulatory capital applications over time.

B. IMPROVED EXAMINATION GUIDANCE

As noted above, most large U.S. banks today have highly disciplined systems for grading the credit quality of individual **financial** instruments within major portions of their credit portfolios (such as large business customers). In combination with other information from banks' internal risk models, these internal grades could provide a basis for developing specific and practical examination guidance to aid examiners in conducting independent assessments of the capital adequacy of large, complex banking organizations.

To give one example, in contrast to the one-sizefits-all Basle standard, a bank's internal capital allocation against a fully funded, unsecured commercial loan will generally vary with the loan's internal credit rating. Typical internal capital allocations often range from 1 percent or less for a grade-1 loan, to 14 percent or more for a grade-6 loan (in a credit rating system with six "pass" grades). Internal economic capital allocations against classified, but not-yet-charged-off, loans may approach 40 percent-not counting any reserves for expected future charge-offs. Examiners could usefully compare a particular bank's actual capital levels (or its allocated capital levels) with the capital levels implied by such a grade-by-grade **analysis** (using as benchmarks the internal capital allocation ratios, by grade, of peer institutions). At a minimum, such a comparison could initiate discussions with the bank on the reliability of its internal approaches to risk measurement and capital allocation. Over time, examination guidance might evolve to encompass additional elements of banks' internal risk models, including analytical tools based on stress-test methodologies. Regardless of the specific details, the development and field testing of examination guidance on the use of internal credit risk models would provide useful insights into the longer term feasibility of an internal models approach to setting formal regulatory capital standards.

More generally, both supervisors and the banking industry would benefit from the development of sound practice guidance on the design, implementation, and application of internal risk models and capital allocation systems. Although important concerns remain, this field has progressed rapidly in recent years, reflecting the growing awareness that effective risk measurement is a critical ingredient to effective risk management. As with trading account VaR models at a similar stage of development, banking supervisors are in a unique position to disseminate information on best practices in the risk measurement arena. In additional to permitting individual banks to compare their practices with those of

peers, such efforts would likely stimulate constructive discussions among supervisors and bankers on ways to improve current risk modeling practices, including model validation procedures.

VI. CONCLUDING REMARKS

The above discussion provides examples by which information from internal credit risk models might be usefully incorporated into regulatory or supervisory capital policies. In view of the modeling concerns described in this summary, incorporating internal credit risk measurement and capital allocation systems into the supervisory and/or regulatory framework will occur neither quickly nor without significant difficulties. Nevertheless, supervisors should not be dissuaded from embarking on such an endeavor. The current one-size-fits-all system of risk-based capital requirements increasingly is inadequate to the task of measuring large bank soundness. Moreover, the process of "patching" regulatory capital "leaks" as they occur appears to be less and less effective in dealing with the challenges posed by ongoing **financial** innovation and regulatory capital arbitrage. Finally, despite difficulties with an internal models approach to bank capital, no alternative long-term solutions have yet emerged.

Footnote:

ENDNOTE

Footnote:

The views expressed in this summary are those of the authors and do not necessarily reflect those of the Federal Reserve System or other members of its staff This paper draws heavily upon information obtained through our participation in an ongoing Federal Reserve System task force that has been reviewing the internal credit risk modeling and capital allocation processes of major U.S. banking organizations. The paper reflects comments from other members of that task force and Federal Reserve staff including Thomas Boemio, Raphael Bostic, Roger Cole, Edward Ettin, Michael Gordy, Diana Hancock, Beverly Hirtle, James Houpt, Myron Kwast, Mark Levonian, Chris Malloy, James Nelson, Thomas Oravez, Patrick Parkinson, and Thomas Williams. In addition, we have benefited greatly from discussions with numerous practitioners in the risk management arena, especially John Drzik of Oliver, Wyman fi Company. We alone, of course, are responsible for any remaining errors.

Reference:

REFERENCES

Reference:

Jones, David, and John Mingo. 1998. "Industry Practices in Credit Risk Modeling and Internal Capital Allocations: Implications for a Models-Based Regulatory Capital Standard." Paper presented at the conference " **Financial** Services at the Crossroads: Capital Regulation in the Twenty-First Century," Federal Reserve Bank of New York, February 26-27.

Author Affiliation:

David Jones is an assistant director and John Mingo a senior adviser in the Division of Research and Statistics of the Board of Governors of the Federal Reserve System.

Special Features: Graphs; Diagrams; References
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 8100 (CN=Financial services industry); 4310 (CN=Regulation)
Descriptors: Studies; Credit risk; Models; Capital requirements; Regulation of **financial** institutions; Safety & soundness
Geographic Names: US

36/9/6 (Item 6 from file: 268)
00348481
Methods for evaluating **value**-at-risk estimates
Lopez, Jose A
Economic Policy Review , v 4 , n 3 , p 119-124 , Oct 1998 **Document Type:** Journal Article **Journal Code:** EPV **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 01768

Abstract:

As implemented in the US, the market risk amendment to the Basle Capital Accord requires that commercial banks with significant trading activity provide their regulators with **value** at risk (VaR) estimates from their own internal models. The VaR estimates will be used to determine the banks' market risk capital requirements. It is important to evaluate the accuracy of VaR estimates from a regulatory perspective. The power of binomial and **interval** forecast evaluation methods can be low against reasonable alternative VaR models. This result does not negate their usefulness, but it does indicate that the inference drawn from this **analysis** has limitations. The proposed loss function method is based on assigning numerical scores to the performance of VaR estimates under a loss function that reflects the concerns of the regulators. This method can provide additional useful information on the accuracy of the VaR estimates. It allows the evaluation to be tailored to specific interests that regulators may have.

Text:

I. CURRENT REGULATORY FRAMEWORK

In August 1996, the U.S. bank regulatory agencies adopted the market risk amendment (MRA) to the 1988 Basle Capital Accord. The MRA, which became effective in January 1998, requires that commercial banks with significant trading activities set aside capital to cover the market risk exposure in their trading accounts. (For further details on the market risk amendment, see Federal Register [1996].) The market risk capital requirements are to be based on the **value**-at-risk (VaR) estimates generated by the banks' own risk management models.

In general, such risk management, or VaR, models forecast the distributions of future **portfolio** returns. To fix notation, let Yt denote the log of **portfolio value** at time t. The k-period-ahead **portfolio** return is Et+ k = Yt

-Yt Conditional on the information available at time t, Et+ k is a **random** variable with distribution fi + k. Thus, VaR model m is characterized by fmt+k, its forecast of the distribution of the k-period-ahead **portfolio** return.

VaR estimates are the most common type of forecast generated by VaR models. A VaR estimate is simply a specified quantile (or critical **value**) of the forecasted fmt + k The VaR estimate at time t derived from model m for a k-period-ahead return, denoted VaRm,(k,a), is the critical **value** that corresponds to the lower a percent tail of fmt + k In other words, VaR estimates are forecasts of the maximum **portfolio** loss that could occur over a given holding period with

a specified confidence level.

Under the "internal models" approach embodied in the MRA, regulatory capital against market risk exposure is based on VaR estimates generated by banks' own VaR models using the standardizing parameters of a ten-day holding period (k = 10) and 99 percent coverage (ox = 1). A bank's market risk capital charge is thus based on its own estimate of the potential loss that would not be exceeded with 1 percent certainty over the subsequent twoweek period. The market risk capital that bank m must hold for time t + 1, denoted MCRmt+ t, is set as the larger of VaRat(10,1) or a multiple of the average of the previous sixty VaRmt(10,1) estimates, that is, (Formula Omitted)

estimates for a one-day holding period (k = 1) and 99 percent coverage, denoted VaR mt(1,1) or simply VaRmt. Smt is a step function that depends on the number of exceptions (that is, occasions when the **portfolio** return Et + I is less than VaRmt) observed over the last 250 trading days. The possible number of exceptions is divided into three zones. Within the green zone of four or fewer exceptions, a VaR model is deemed "acceptably accurate," and Smt remains at its minimum **value** of three. Within the yellow zone of five to nine exceptions, Smt increases incrementally with the number of exceptions. Within the red zone of ten or more exceptions, the VaR model is deemed to be "inaccurate," and Smt increases to its maximum **value** of four.

II. ALTERNATIVE EVALUATION METHODS Given the obvious importance of VaR estimates to banks and now their regulators, evaluating the accuracy of the models underlying them is a necessary exercise. To date, two hypothesis-testing methods for evaluating VaR estimates have been proposed: the binomial method, currently the quantitative standard embodied in the MRA, and the **interval** forecast method proposed by Christoffersen (forthcoming). For these tests, the null hypothesis is that the VaR estimates in question exhibit a specified property characteristic of accurate VaR estimates. If the null hypothesis is rejected, the VaR estimates do not exhibit the specified property, and the underlying VaR model can be said to be "inaccurate." If the null hypothesis is not rejected, then the model can be said to be "acceptably accurate." However, for these evaluation methods, as with any hypothesis test, a key issue is their statistical power, that is, their ability to reject the null hypothesis when it is incorrect. If the hypothesis tests exhibit low power, then the probability of misclassifying an inaccurate VaR model as "acceptably accurate" will be high. This paper examines the power of these tests within the context of a simulation exercise. In addition, an alternative evaluation method that is not based on a hypothesis-testing framework, but instead uses standard forecast evaluation techniques, is proposed. That is, the accuracy of VaR estimates is gauged by how well they minimize a loss function that represents the regulators' concerns. Although statistical power is not relevant for this evaluation method, the related issues of comparative accuracy and model misclassification are examined within the context of a simulation exercise. The simulation results are presented below, after the three evaluation methods are described. (See Lopez [1998 for a more complete discussion.)

EVALUATION OF VAR ESTIMATES BASED ON THE BINOMIAL DISTRIBUTION

Under the MRA, banks will report their VaR estimates to their regulators, who observe when actual **portfolio** losses exceed these estimates. As discussed by Kupiec (1995), assuming that the VaR estimates are accurate, such exceptions can be modeled as independent draws from a binomial distribution with a probability of occurrence equal to 1 percent. Accurate VaR estimates should exhibit the property that their unconditional coverage a* = x/250, where x is the number of exceptions, equals 1 percent. Since the probability of observing x exceptions in a sample of size 250 under the null hypothesis is (Formula Omitted)

(Formula Omitted)

Loss Function That Addresses the Magnitude of the Exceptions As noted by the Basle Committee on Banking Supervision (1996), the magnitude as well

as the number of exceptions are a matter of regulatory concern. This concern can be readily incorporated into a loss function by introducing a magnitude term. Although several are possible, a quadratic term is used here, such that (Formula Omitted)

eight VaR models evaluated and is designated as the "true" model, or model 1.

The next three alternative models are homoskedastic VaR models. Model 2 is simply the standard normal distribution, and model 3 is the normal distribution with a variance of 1. Model 4 is the t-distribution with six degrees of freedom, which has fatter tails than the normal distribution and an unconditional variance of 1 .

The next three models are heteroskedastic VaR models. For models 5 and 6, the underlying distribution is the normal distribution, and hmt + 1 evolves over time as an exponentially weighted moving average of past squared returns, that is, (Formula Omitted)

This type of VaR model, which is used in the well-known RiskMetrics calculations (see J.P. Morgan 1996)), is calibrated here by setting k equal to 0.94 and 0.99 for models 5 and 6, respectively. Model 7 has the same variance dynamics as the true model, but instead of using the normal distribution, it uses the t-distribution with six degrees of freedom. Model 8 is the VaR model based on **historical** simulation using 500 observations, that is, using the past 500 observed returns, the a percent VaR estimate is observation number 5*a of the sorted returns. In the table, panel A presents the power **analysis** of the hypothesis-testing methods. The simulation results indicate that the hypothesis-testing methods can have relatively low power and thus a relatively high probability of misclassifying inaccurate VaR estimates as "acceptably accurate." Specifically, the tests have low power against the calibrated normal models (models 5 and 6) since their smoothed variances are quite similar to the true GARCH variances. The power against the homoskedastic alternatives is quite low as well.

For the proposed loss function method, the simulation results indicate that the degree of model misclassification generally mirrors that of the other methods, that is, this method has a low-to-moderate ability to distinguish between the true and alternative VaR models. However, in certain cases, it provides additional useful information on the accuracy of the VaR estimates under the defined loss function. For example, note that the magnitude loss function is relatively more correct in classifying VaR estimates than the binomial loss function. This result is not surprising given that it incorporates the additional information on the magnitude of the exceptions into the evaluation. The ability to use such additional information, as well as the flexibility with respect to the specification of the loss function, makes a reasonable case for the use of the loss function method in the regulatory evaluation of VaR estimates. (Formula Omitted)

(Formula Omitted)

(Formula Omitted)

IV CONCLUSION As implemented in the United States, the market risk amendment to the Basle Capital Accord requires that commercial banks with significant trading activity provide their regulators with VaR estimates from their own internal models. The VaR estimates will be used to determine the banks' market risk capital requirements. This development clearly indicates the importance of evaluating the accuracy of VaR estimates from a regulatory perspective. The binomial and **interval** forecast evaluation methods are based on a hypothesis-testing framework and are used to test the null hypothesis that the reported VaR estimates are "acceptably accurate," where accuracy is defined by the test conducted. As shown in the simulation exercise, the power of these tests can be low against reasonable alternative VaR models. This result does not negate their usefulness, but it does indicate that the inference drawn from this **analysis** has limitations.

(Illustration Omitted)

Captioned as: SIMULATION RESULTS FOR GARCH(1,1)-NORMAL DGP Units: percent

The proposed loss function method is based on assigning numerical scores to the performance of the VaR estimates under a loss function that reflects the concerns of the regulators. As shown in the simulation exercise, this method can provide additional useful information on the accuracy of the VaR estimates. Furthermore, it allows the evaluation to be tailored to specific interests that regulators may have, such as the magnitude of the observed exceptions. Since these methods provide complementary information, all three could be useful in the regulatory evaluation of VaR estimates.

Reference:

REFERENCES

Reference:

Basle Committee on Banking Supervision. 1996. "Supervisory Framework for the Use of 'Backtesting' in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." Manuscript, Bank for International Settlements.

Christoffersen, P. F. Forthcoming. "Evaluating **Interval** Forecasts." INTERNATIONAL ECONOMIC REVIEW.

Federal Register. 1996. "Risk-Based Capital Standards: Market Risk." Vol. 61: 47357-78.

Reference:

J.P. Morgan. 1996. RISKMETRICS TECHNICAL DOCUMENT. 4th ed. New York: J.P. Morgan.

Kupiec, P. 1995. "Techniques for Verifying the Accuracy of Risk Measurement Models." JOURNAL OF DERIVATIVES 3: 73-84. Lopez, J. A. 1998. "Methods for Evaluating **Value**-at-Risk Estimates." Federal Reserve Bank of New York Research Paper no. 9802.

Author Affiliation:

Jose A. Lopez, formerly an economist at the Federal Reserve Bank of New York, is now an economist at the Federal Reserve Bank of San Francisco.

Special Features: Charts; Graphs; Equations; References
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 8100 (CN=Financial services industry); 3100 (CN=Capital & debt management)
Descriptors: Studies; Risk management; Regulation of **financial** institutions; Capital requirements
Geographic Names: US

36/9/7 (Item 7 from file: 268)
00329067
Tools for measuring interest-rate risk
Taylor, Jeremy F
Bank Accounting & Finance , v 11 , n 2 , p 7-13 , Winter 1997/1998 **Document Type:** Journal Article
ISSN: 0894-3958 **Journal Code:** ABAF **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER: L
Word Count: 02871

Abstract:

Asset/liability managers have made substantial progress in dealing with

interest rate-risk issues. By understanding the different methods of measuring risk, risk managers will be better able to manage interest-rate and other risks. To calculate duration, duration is adjusted for noncontinuous interest compounds; the adjustment is greatest for longer maturity, higher-yielding instruments. Total return and return on risk-adjusted capital approaches began to reflect the complexities, but not the uncertainties of balance sheets. It became possible to model uncertainties using option-adjusted spread pricing techniques that expressed future choices in terms of option theory by linking optionality with probabilistic Monte Carlo rate simulation. Modeling risk occurs because at least 3 distinct perspectives on risk can be found, each internally coherent but at cross-purposes with the others.

Text:

Headnote:

For asset/liability managers, the risk of the models they choose to manage rate risk is, itself, not inconsiderable. This article reviews the different techniques for measuring interest-rate risk and shows how they vary in goals and emphasis.

In the past decade, three major theoretical constructs-modified duration, **value** at risk, and option-adjusted spread (OAS) pricinghave successively advanced risk-management metrics from a relatively crude, highly structured, forecasting-oriented set of simulation methods to a subtle, sophisticated, and probabilistic representation of risk uncertainties. Projected cash flows of net-interest income have given way to capitalized measures of risk. Likewise, earlier conceptualizations of a balance sheet as an entity with a defined internal structure that could be forecast forward on a linear basis have given way to the realization that to forecast the performance of balance sheets and off-balance-sheet instruments, one must include choices yet to be exercised. These choices can affect instruments of every conceivable form. They include prepayments, activation of lease residuals, early withdrawals, discretion in deposit pricing, and potentials for disintermediation. The choices involve myriad puts and calls. Sometimes these are "virtual," as in an early deposit withdrawal, sometimes actual, as in callable bonds or balloon loans.

Asset/liability managers have made substantial progress in dealing with rate-risk issues. Yet the advances in technique at the boundaries of the rate-risk-management field have not been mirrored by changes in understanding throughout the population of riskmanagement practitioners, as shown by recent losses discovered when mismodeled portfolios have been remarked to market. In March 1997, for example, National Westminster Bank PLC acknowledged a $139 million modeling error, and Bank of Tokyo took an $83 million writeoff. According to Capital Market Risk Advisors, "model risk" has been a "major cause" of derivatives losses totaling $1.9 billion.1

It is absurd that risk-management innovations should themselves have created new risks. By understanding the different methods of measuring risk, risk managers will be better able to manage interest-rate and other risks. This article will review and evaluate rate-risk techniques to put them in context and describe their purposes.

Capitalizing Rate Risk

The shift from net-interest-income measures of rate risk toward capitalization methods must be credited to those who had the greatest interest in protecting bank capital, that is, the banking regulators. To establish standard measures of institutional rate risk and to protect depository insurance, as required by the FDIC Improvement Act of 1991 (FDICIA), the regulators have moved toward modified duration mathematics to

relate changes in the yield environment to changes in **portfolio value**.

Duration is the timed average cash flow of a bond weighted by the present values of those cash flows. Conceptually, duration is the future point in time at which, on average, onehalf of an investment has been realized in present-**value** terms. Duration increases at a decreasing rate as maturity is extended; the decrease in change is a function of the exponential decrease in the present **value** of the final payments, which include repayment of the face investment amount.

Low coupons translate into longer durations, as the difference between instrument price and face investment repayment at maturity assumes greater weight over the interest cash flows. Long maturities increase this effect as the discount on face repayment magnifies through time. Factors influencing bond duration are summarized in Exhibit 1.(2)

Modified duration, in turn, is Duration/(1 + Internal rate of return/Number of compounding periods per year). To calculate modified duration, duration is adjusted for noncontinuous interest compounds; the adjustment is greatest for longer maturity, higher-yielding instruments. Hence, duration can be applied across all instrument types, a prerequisite to using it in a standard regulatory framework.3 Modified duration acts as a multiplier and estimates an instrument's price sensitivity to changes in market yield, reflecting the price/yield trade-off as a linear first derivative (Exhibit 2).

Error resulting from linearity (as shown in Exhibit 2) eventually caused the FDICIA regulatory proposal to prove impracticable. Duration is accurate only for small, instantaneous, parallel shifts in the yield curve; therefore, it is not appropriate to rate-shock tests. The basic regulatory purpose was to subject balance sheets to stress tests (such as 400 basis point shocks) that would uncover institutions that posed a risk to the insurance fund. After lengthy review, the regulators abandoned the duration measure.

The yield curve is a series of infinite discontinuous straight-line tangents (Exhibit 3). For continuous curvilinear yield/price relationships, it is necessary to include convexity, the second derivative of the price/yield function. Assuming parallel yield-curve shifts, the price volatility multiplier for falling yields is modified duration plus the convexity factor, for rising yields, modified duration minus the convexity factor. Noncallable bonds have positive convexity, which is greatest in long bonds under large yield changes (Exhibit 4).

Callable bonds have negative convexity over specific ranges in interest rates close to the call price. As yields fall, bond prices increase at a decreasing rate and peak slightly above the call price, therefore declining as the probability of call approaches certainty (Exhibit 5).

(Table Omitted)

Captioned as: Exhibit 1

(Graph Omitted)

Captioned as: Exhibit 2

(Graph Omitted)

Captioned as: Exhibit 3

Without incorporating convexity, a regulatory measure of duration would not have stood the test of real-world relevance. So, the regulators stopped trying to find a consistent, numerical measure and instead emphasized an institution's risk-management "process."4 The regulators would assess risk on a case-by-case basis, and competing modeling and measurement standards were permitted.

Option-Adjusted Spread Pricing, Linear Path Space, and **Value** at Risk

(Chart Omitted)

Captioned as: Exhibit 4

(Chart Omitted)

Captioned as: Exhibit 5

When assessing rate risk, the interests of regulators on the one hand and the **financial** institutions on the other are not the same. The latter do not care, except abstractly, about the prospects of going out of business. (For more, see "Economic Interest-Rate Risk and the Bank as a Going Concern" on page 25 in this issue.) **Financial** institutions care about profit margins and protecting and enhancing the income streams of an ongoing enterprise. So, as the regulators were looking at duration, the industry developed different techniques of risk measurement.

When assessing rate risk, the interests of regulators on the one hand and the **financial** institutions on the other are not the same.

Interest-rate-risk managers, after developing simulation tools in the late 1980s, were most concerned with resolving uncertainties surrounding cash flows. Risk managers needed to model highly fluid instrument terms generated by the disaggregation of **financial** risks, specifically, derivatives.

Total return and return on risk-adjusted capital (RAROC) approaches began to reflect the complexities, but not the uncertainties, of balance sheets. It became possible to model uncertainties using option-adjusted spread pricing techniques that expressed future choices in terms of option theory by linking optionality with probabilistic Monte Carlo rate simulation.

Option-adjusted spread

Option-adjusted spread pricing models use simulations to evaluate the impact of embedded options on the **value** of an institution's assets, liabilities, and off-balance-sheet instruments (Exhibit 6). Important options in assets include prepayments, loan calls, exercisable lease puts, caps, floors, and collars. Options in liabilities include disintermediations of deposits, betas on core-deposit repricing, and early withdrawals of term liabilities, such as certificates of deposit. Off-balance-sheet options include swap calls at the end of lockout periods.

The steps to model option-adjusted pricing for assets follow:

*Create **random** rate paths.

*Set prepayment rates and other options.

* Calculate remaining principal and annual cash flow by path.

* Discount the simulated cash flow to present **value** using simulated Treasury rates.

*Express the resulting **value** as a spread over the Treasury rate.

*Calculate the present **value** of all cash flows across all scenarios, or optionadjusted spread.

To calculate option-adjusted pricing for liabilities, prepayment rates and other options are replaced with betas, which are varied depending on whether rates are rising or falling. By creating **random** rate scenarios, one can establish average spreads to Treasury across scenarios.

Therefore, the option-adjusted model price of an instrument can be compared to prevailing market price. Instruments that price over the market have positive option-adjusted spreads. On a institutional basis, therefore, OAS **value** can be compared to market **value**. As institutions prudently manage their characteristics over time, positive OAS **value** can result, which is reflected in purchase premiums. As rates increase, for example, institutions can lag deposit repricing betas while prepayments fall. Transactionally, institutions should purchase instruments that carry positive OASmarket spreads. For hedging, the normal rule is to offset current option-adjusted durations.

Important risk derivatives are drawn from option pricing theory, including the following:

* Price delta: price change.

*Price gamma: second derivative of price change, expressing speed of change. Increased speed will be exemplified by atthe-money options close to expiration or bonds with increasing probability of call.

* Volatility delta (Vega): change in option premium given a 1% change in attributed volatility. Vega is greatest at-the-money with long

expiration.

*Time delta (Theta): decline in option premium given a one-day passage of time. Option time **value** falls more quickly as time-to-expire diminishes.

* Interest-rate delta (Rho): change in option premium from change in the risk-free rate.

Linear path space

(Table Omitted)

Captioned as: Exhibit 6

The practical shortcoming of Monte Carlo simulations is the time it takes to generate the many, many scenarios necessary to cover the range of rate possibilities. Linear path space mechanics aim to exclude the repeated generation of high-probability scenarios under fully randomized techniques in order to create lowprobability forecasts (path 3 in Exhibit 7). Monte Carlo simulations involve numerous repetitions of paths toward the middle of the binomial lattice in order to occasionally create outer lattice/low-probability rate scenarios. Linear path space selectively excludes paths already generated in order to cover useful rate ranges more efficiently.

Value at risk

The contrast between the regulatory perspective, which concentrates on worst-case outcomes, and managerial perspectives, which seeks assistance in everyday decision making, is clear in the differences between the concepts of "**value** at risk" and "economic **value** of equity" (EVE). Regulators have tied economic **value** of equity explicitly to stress-test criteria. **Value** at risk, on the other hand, is defined as the maximum possible loss within a given confidence **interval** over an orderly liquidation period (Exhibit 8).

The basic elements in a **value**-at-risk **analysis** are sensitivity to market factors, volatility of market factors, the holding period, the selected confidence **interval**, and correlation among market factors. To model **value** at risk, one may use various rate scenario approaches, including delta normal correlation matrices and **historically** based scenarios. The most exhaustive approaches, which are most suitable to a risk manager's needs, are Monte Carlo or linear path space simulations.

Complexity and the Limits of Market **Value**

Uncertainty is not the same thing as complexity. Option-adjusted spread models aim to address uncertainty rather than complexity. Complexities can arise from instrument structure, completely aside from any optionality factors. A **portfolio** of leveraged leases, for example, will have a highly differentiated set of cash flows, some of them internal, as lease amortizations and residual agreements combine with tax deferrals and equity positions in the form of retained earnings. In such a case, total return is highly certain, except for event risks as presented by the possibility of changing tax laws and risks aside from interest-rate risk, such as credit risk.

Option-adjusted spread models aim to address uncertainty rather than complexity.

(Chart Omitted)

Captioned as: Exhibit 7

(Table Omitted)

Captioned as: Exhibit 8

In modeling such instruments, there is great need for technical understanding of the formal returns embodied within the underlying transaction. So, management of instruments that are laden with interest-rate risk increasingly involves other sorts of expertise. The boundaries between interest-rate-risk management and previously unconnected fields, such as tax accounting, grow less distinct as **financial** arrangements become more complex.

At the same time, option-adjusted spread models carry the implication that they may be more accurate than the market. The presumption is that an

OAS model will uncover "market mispricing...for example, the market over-compensation on the call risk in a mortgage-backed security."5 The implication that OAS pricing may be more accurate than the market creates an interesting paradox for the treatment of OAS models. They may be right when the market is wrong or wrong when the market is wrong. In either case, however, the portfolios they **value** must be marked to market, and institutions must take the write-off when the market dictates.

When a **portfolio** ought to be (re)marked to market is not as self-evident, or categorical, as one might presume. The difference, it must be remembered, is between market price and OAS price (which is a matter of fact) not between a manifest flaw and a debatable one (which is a matter of assumption).

This is not intended to undermine the notion of modeling risk, only to show that it is subtle. Sometimes market **value** is not of primary relevance. One can, for example, provide a calculated market **value** of a **portfolio**, as well as a full accounting schedule for amortization of fair-**value** adjustments, using an OAS model. The fact is, however, that the prices offered for such a **portfolio** will vary just as in any other sale. Bid price is a function of perceived strategic opportunity. One might hope to reduce such an opportunity to mathematical terms, such as hurdle rates or earnings per share. But in the final **analysis**, **value** is in the eye of the beholder, and those who behold it will pay a higher price than those who don't.

Modeling Risk

Modeling risk occurs because at least three distinct perspectives on risk can be found, each internally coherent but at cross-purposes with the others: Regulators seek the avoidance of catastrophic institutional loss. Risk managers seek to incorporate uncertainty into modeling techniques. Executive managers seek a simple, straightforward risk-management language.

We can compare regulators' and risk managers' points of view by looking at the way each might evaluate the adequacy of risk-management software. In the regulatory view, **financial** institutions fall into one of three categories:

*Simple, short-term, stable.

*Complex, longer-term, volatile.

* The middle ground, into which all but the largest or the most conservative institutions fall.

Regulators would define the basic analytical need of each bank by its most predominant risk characteristic (Exhibit 9). Rate-scenario custom building facilities

O Severe up/down stress test

O Yield-curve rotation

O Individual rate manipulation

O Path-dependent discounted cash flow

OBalance-building facilities

O Rate/balance interdependence

O Maximum balance flexibility

O Multiple balance sheet generation

O Individual balance simulation

OFlexibility of modeling

O Variable chart of accounts

O Mix and match rate/balance connection

O Strategic business-unit hierarchy

OForecasting credibility

O Accurate accrual

(Table Omitted)

Captioned as: Exhibit 9

(Table Omitted)

Captioned as: Exhibit 10

O Translation ease to custom bank report format

O Variance capability by budget matrices

OProfitability measurement
O Credible funds transfer
O Plausible relationship pricing
O Statistical profitability function
OData management
O Ease of upload/download linkage
O Compatibility with existing system applications
O Compatibility with existing reporting segmentations

More complex or volatile institutions may require stochastic OAS modeling.

Regulators would call the following steps critical in evaluating the software:

*situation **analysis**-current needs/future requirements;
*system optimization profile (of available software);
* cost-effectiveness study;
*resource management (for implementation);
* critique vendor presentations of software suitability;
* fit with statement of purpose: regulatory requirement vs. strategic modeling.

Regulatory concerns aim to isolate a highly specific outcome: disaster. Managers, on the other hand, look over many possible outcomes under realistic circumstances in search of opportunity.

Senior management's view

The oversight function provided by senior management of those who perform the professional task of measuring rate risk has been lost. The complexity of risk management has made its language inaccessible to the senior managers it is supposed to inform. It is hard to believe, for instance, that senior managers know (or care) about the vega of an instrument. The present situation should not be allowed to persist. Whether in the everyday strategic management of an institution's business or in the satisfying of regulatory mandates to protect the deposit insurance fund, the mission of interest-rate-risk managers should explain risk, not make it unintelligible. Exhibit 10 presents specifications for an interest-rate-risk model suitable for a mediumsized, stable, moderately complex **financial** institution. The field has made great progress in incorporating important aspects of risk into the measurement of the impact of changes in interest rates. Now, it is time to share these improvements with those who are not specialists.

Footnote:

Notes

1 Capital Market Risk Advisors of New York, quoted in Suzanne McGee, "Bank of Tokyo Blames Loss on Bad Model" Wall Street Journal, March 28, 1997, sec. A, p. 11. 2For a discussion of how to calculate duration, see Livingston G. Douglas, Yield Curve **Analysis** (New York: Simon 8c Schuster, 1988): 158.

3The final version of the proposed regulatory rule is found in the Federal Register, August 2, 1995, p. 39490. "For a review of this rule see Kevin P Riley and Jeremy F. Taylor, "Stealth Regulation Files Over Interest Rate Risk" ABA Banking Journal (October 1996): 93. 5 Sirri S. Ayaydin, Charles A. Richard III, and Stephen Rigsbee, "Total Return Approach to Asset/Liability Management Using Option-based Models," Asset/Liability Management, Frank J. Fabozzi and Atsuo Konisi, eds. (Chicago: Probus Publishing, 1991): 165.

Author Affiliation:

Jeremy F. Taylor has written many articles on asset/liability management and four books on banking, including The Prudent Management of Modern U.S. Banks (Westport, Conn.: Quorum Books, 1994).

Special Features: Graphs; Charts; References
Descriptors: Interest rate risk; Asset liability management; Risk management; **Financial** institutions; Models
Geographic Names: US

36/9/8 (Item 8 from file: 268)
00322220
Bank deposits and credit as sources of systemic risk
Eisenbeis, Robert A
Economic Review (Federal Reserve Bank of Atlanta) , v 82 , n 3 , p 4-19 , Third Quarter 1997
CODEN: ECRWDA **Document Type:** Journal Article **ISSN:** 0732-1813 **Journal Code:** ECR **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 09213

Abstract:

An investigation of various theories of **financial** panics and crises, which places particular emphasis on the links between credit and deposits, suggests that panics are neither perfectly predictable nor **random** events. Information asymmetries about banks' ability to liquefy deposits were apparently a major contributing factor to banking panics in the past. In addition, **financial** crises do not seem to have been primary causal agents of recessions. The **analysis** also suggests that government policies can affect the likelihood of a **financial** crisis as well as play a role in its solution. The dynamics of crises may differ significantly in the future given recent, rapidly developing changes in the US and world **financial** system.

Text:

THERE ARE MANY DIFFERENT WAYS TO DEFINE A **FINANCIAL** CRISIS. INDEED, THE ECONOMICS AND **FINANCE** LITERATURE IS FILLED WITH TERMS LIKE PANIC, **FINANCIAL** CRISIS, RUNS, SYSTEMIC CRISIS, OR CONTAGION.1 THERE IS IN FACT LITTLE AGREEMENT ON EVEN THE RUDIMENTARY DEFINITIONS OF A **FINANCIAL** CRISIS, THE SEQUENCE OF EVENTS CONSTITUTING A CRISIS, OR THE CAUSES OF THESE EVENTS.

The professional discussion divides itself into two broad categories. Macroeconomists typically are concerned with explaining business cycle fluctuations and determining when a recession will degenerate into a depression.2 They are equally interested in the **financial** system's role as a propagator of this process because most depressions have been accompanied by serious disruptions in the **financial** system, including banking failures and panics. Eichengreen and Portes, for example, define a **financial** crisis as "a disturbance to **financial** markets, associated typically with falling asset prices and insolvency among debtors and intermediaries, which ramifies through the **financial** system, disrupting the market's capacity to allocate capital within the economy. . . . Our definition implies a distinction between generalized **financial** crisis on the one hand

and bank failures, debt defaults and foreign-exchange market disturbances on the other" (1991, 10).

Financial economists examine the micro behavior of market participants to explain disruptions in **financial** markets (see Diamond and Dybvig 1983; Chari and Jagannathan 1984). They have tended to focus on banking panics and runs and the reasons depositors withdraw funds rather than on the macro consequences for employment and output in the real economy per se.

While differing in their emphases, the micro and macro approaches to analyzing **financial** stability share several themes. The first focuses on alternative explanations for why a crisis occurs. One prominent thesis argues that the **financial** system is inherently unstable and is therefore vulnerable to **random** shocks. Shocks simultaneously cause market participants to lose confidence in the system and exchange their bank deposits for currency. Others believe that such herd behavior cannot be explained solely by shocks that, like animal spirits, **randomly** induce depositors to run from bank deposits to currency. They offer more behaviorally oriented explanations and models, the most prevalent being models based on the existence of information asymmetries between borrowers and lenders. These models attempt to show that it is sometimes rational for depositors to attempt to withdraw their funds in such a way that it creates a run on the banking system.

Most of the **analysis** in the **random** shock and information asymmetries models concentrates on aggregate behavior, assuming essentially that all market actorsboth depositors and institutions-are identical. It does not admit differences among depositors and institutions or even the presence of more than one institution in the **financial** system. When the **analysis** recognizes more realistic features of market and **financial** structure, researchers are better able to examine the process by which a shock or problem in one part of the country or sector of the economy is transmitted to other sectors or the system as a whole. These transmission models, representing the second main theme in the literature, have not been the focus of much empirical work and tend to be relatively undeveloped.

The third area investigates the causes of **financial** crises and their impact on the real economy. For example, do **financial** crises cause declines in real economic output, or are they instead manifestations of deeper problems in the real economy? What are the channels of transmission? Do deposit runs cause liquidity problems, which in turn induce contractions in lending, thereby affecting real output and production? The final area of **analysis** examines the role of government policiesboth macro and micro-in generating **financial** crises as well as lessening their potential severity.

The remainder of this article explores these issues in more depth. The discussion gives particular attention to the possible linkages between deposits and credit availability as the transmission mechanism for crises since runs on deposits and payments system disruptions are believed to be transmitted to the real economy through a credit channel.

Random Shocks and Inherent **Financial** Instability

At the macroeconomic level, models such as those proposed by Minsky (1982) and Kindleberger t1978) embody the claim that the banking system is inherently unstable. Minsky argued that a capitalist economy, and especially its banking system, is inherently unstable. Furthermore, this instability is endogenous, originating within the system itself. He defined instability as "a process in which rapid and accelerating changes in the prices of assets (both **financial** and capital) take place relative to the prices of current output" (1982, 13).

Simply stated, Minsky assumed that during relatively stable times firms engage in balanced **financing**, by which he meant that cash flows are sufficient to cover principal and interest payments. However, as the economy grows and enters the expansion phase of the business cycle, firms begin to reach for profits, presumably because of management's

preference for shortterm gains. Firms start to leverage up, and banks, in particular, begin to shorten the maturity structure of their liabilities relative to their assets. Expanding returns by funding long-term investments with shortterm borrowing is driven by the desire to take advantage of an upward-sloping term structure with long-term interest rates exceeding short-term rates.3 This period of leveraging, which Minsky labels a period of speculative **finance**, is still one of relative stability.

Cash flows from investment are still sufficient to cover principal payments as debts. This speculation ultimately degenerates into what Minsky calls a period of Ponzi **finance**, in which cash flows cover neither principal nor interest payments. Debt refunding requires new debt issuance, the proceeds of which are used to cover required interest and principal debt payments. During this period, an exogenous shock will result in a collapse of both the **financial** system and the real economy. The shock, which can come from many different sources, serves as the trigger for collapse. Minsky was silent on the exact mechanisms by which this happens.

Commenting on Kindleberger's (1978) similar view, Schwartz observes that "those who regard banks as inherently unstable assume no connection between monetary policy and the price conditions under which economic agents make decisions. Proponents of inherent instability see a recurring **historical** pattern in which many bankers abandon conservative standards of asset management during business expansions only to be caught short when booms collapse. For them instability resides in economic agents. Benevolent government then comes to the rescue. This is the central thesis offered by Charles P. Kindleberger in his 1978 book" (1986, 11).

Minsky puts forth certain stylized facts that would be observed, although they are not the outcome from any specified model.4 The first is that, during an expansion, credit expands at rates that exceed the growth of income or the capital stock. Second, interest rates and nominal asset prices must be increasing at a rapid rate. Third, debt maturities must become shorter, and, fourth, some exogenous shock must occur to cause a change in expectations. Finally, governments must fail to intervene in ways that cushion any asset reevaluations accompanying any changes in expectations.5

Bernanke and James (1991) suggest a different view of the causal relationship. For them, a precipitating force that could lead to a **financial** collapse is a deflation. Deflation adversely affects credit quality by reducing borrower equity cushions. When companies finally default, intermediaries become owners of illiquid, real assets. To reliquefy their balance sheets, banks are induced to reduce lending and call in loans. Those banks that are unable to reliquefy fail, and, by implication, deposits will be destroyed. In this scenario, credit problems lead to a reduction in bank deposits, contracting the money supply.

The chief distinction from the picture Minsky and Kindleberger paint is that Bernanke and James see banks as passive bystanders in the process. They are not required to take on more risk, nor do they have to misprice risk or adjust their balance sheets to take on more interest rate or maturity risk. The model also suggests that crises occur only during and after an exogenous shock has induced a deflation. Bernanke and James are careful to argue, however, that while deflation is a necessary condition for a crisis to occur, it is not sufficient. They highlight several aspects of banking structure that, if present, also help increase the likelihood that **financial** institutions would experience a crisis. These include (a) lack of branch banking, (b) universal banking and the commingling of banking and commerce, and (c) funding though short-term, foreign deposits. Thus, banking and **financial** structure can either mitigate or accentuate the likelihood that a **financial** crisis will result during a deflationary period.

Unlike the macroeconomists' models discussed, the **random** shock models of the **financial** economists, most closely

associated with Diamond and Dybvig (1983), look more deeply at the structure of the deposit contract and the process by which it is redeemed.6 Because deposits are payable upon demand at par, they offer depositors nearly costless liquidity, provided that not all depositors wish to withdraw their funds at the same time. With sequential servicing, in which depositors are treated on a first-come, first-served basis, depositors, especially if they are geographically dispersed, rationally know that not everyone can withdraw simultaneously. If bank loans are inherently not marketable, or cannot be easily liquefied, then at the first hint of potential trouble, it is rational for depositors to step to the head of the line rather than incur costs to determine whether and exactly when deposits will be paid.

These micro **random** shock models pay no particular attention to the source, or nature, of the **random** shock that causes depositors to line up. Depositors just decide to run, and once they do, all depositors run. These models also do not consider the credit side of the balance sheet as a factor in crises, other than the fact that loans are less liquid than deposits so that banks cannot pay all claims in currency. Nonetheless, they make it easy to see that shocks affect depositors' willingness to hold bank deposits, and, when that willingness is reduced, a contraction in credit follows as loans must be liquidated to meet the deposit-redemption demand.

The Diamond and Dybvig model approximates the situation that prevailed in early U.S. history. Individual banks issued their own bank notes to the public, promising to redeem these notes at par for specie.7 Since note issues typically were not backed 100 percent by specie, periodic liquidity problems arose whenever noteholders became concerned that a bank might not be able to honor its redemption commitment and suspend convertibility of deposits into specie. Runs on individual banks and the system sometimes occurred, and these resulted, albeit infrequently, in cumulative contractions in the money stock.8 Suspension of convertibility of deposits into specie was a common way for early banks to deal with temporary liquidity problems. It often resulted, however, in a decline in purchasing power since the **value** of deposits declined. By shifting the cost of nonconvertability at least temporarily to the creditors (depositors) of the bank, they gave all liability holders an important incentive to worry about bank solvency.

Diamond and Dybvig (1983) investigate the suspension of convertibility as one equilibrium solution to the problem of runs, but they do not consider the price level effects or how the costs of suspension of convertibility are distributed because their model has only a consumption good and no currency. Another weakness of their model is that there is only one bank in the system, and hence runs are on the banking system as a whole and involve flights to the currency rather than runs on one of many banks in the system.9

For these early banks, avoidance of runs meant maintaining public confidence. Depositors needed to believe that the institution could convert notes into specie in sufficient amounts and would not need to suspend convertibility.lo Indeed, the first forms of public regulation designed to deal with the problems involving suspension of convertibility imposed reserve requirements specifying permissible ratios of notes to specie. The regulations sought to assure public confidence by requiring banks to engage in minimal maturity intermediation, maintain sufficient specie reserves, and have adequate capital and liquidity.ll

Information Asymmetries Models

The micro **random** shock models have been less than satisfying, both because they appear generally inconsistent with economic events as will be discussed in the next section, and because many economists find it hard to believe that people **randomly** decide to run without some just cause rooted in economics. Recent modeling efforts have applied concepts of information asymmetries to derive conditions that might make it rational for depositors to engage in runs on banks. Under the

information asymmetries models, banks are viewed as being "opaque" to depositors and thus costly for depositors to monitor. With imperfect and costly information, a type of Akerlof (1979) lemons model applies in which depositors have a great deal of difficulty distinguishing between healthy and unhealthy banks. Any shock or news event that might induce depositors to reassess their bank's riskiness (in combination with the sequential servicing constraint) will cause depositors to assume that all banks are riskier than previously believed. Under these circumstances, it is more rational for depositors to withdraw funds than to seek out and evaluate costly information or risk losing their funds by not withdrawing. In these models, as in the micro **random** shock models, the source of the shock is not specified, in that no particular cause is suggested for a failure. But usually it is hypothesized that the shock originates in credit markets and in releases of relevant news about bank asset quality. The model's predictions are consistent with the view that shocks are more likely to result from disturbances in the real sector than from the default of a single borrower.

Macroeconomists have articulated a form of this same asymmetric information hypothesis in attempting to counter the inherent instability arguments. As Schwartz describes it, "a widely held belief in the United States and the world **financial** community is that the default of major debtors–whether companies or municipalities or sovereign countries–could lead to bank failures that would precipitate a **financial** crisis. . . . A **financial** crisis is fuelled by fears that means of payment will be unobtainable at any price and, in a fractionalreserve banking system, leads to a scramble for highpowered money. It is precipitated by actions of the public that suddenly squeeze the reserves of the banking system. In a futile attempt to restore reserves, the banks may call loans, refuse to roll over existing loans, or resort to selling assets.... The essence of a **financial** crisis is that it is short lived, ending with a slackening of the public's demand for additional currency" (1986, 11).12

Under this scenario, a banking crisis is precipitated by the failure of a major debtor, which induces a sudden shift in the public's demand for currency. In turn, banks scramble for reserve assets by curtailing lending and selling assets. By implication the decline in lending and refusal to roll over existing credits leads to a decline in economic output. The process becomes systemic in that problems in one or several major creditors raise questions about the quality of bank assets in general and induce the public to switch to holding currency.

The hypothesis implies a direct linkage between increased demand for money and the availability of credit–and hence the ability to **finance** and maintain the real economy. The information the public perceives is not assumed to be bank specific; instead, it is the fact that the information concerns the quality of banking assets in the aggregate that increases the private sector's demand for currency relative to deposits. The channel envisioned in this scenario results in banks calling in loans and building up liquidity to meet the public's desire for currency.

Empirically, three elements are necessary for this view to hold. First, there must be a credit-related shock that affects the public's desire to hold currency relative to deposits. Second, this shock must induce a liquidation of deposits for currency by the public. Third, bank credit must contract.

There are several important differences between the various **random** shock models and asymmetric information models. First, Minsky's **random** shock model includes leveraging up of both bank and corporate balance sheets across the board, and, furthermore, it does not require an inflationary environment. Second, the collapse that results is not driven by runs forcing institutions to liquefy balance sheets to meet deposit withdrawals. Third, under this type of model **financial** institutions accommodate the leveraging up of balance

sheets by underpricing credit risk. They also take on more interest rate and maturity risk by shortening the maturity structure of liabilities relative to assets. Fourth, no interdependence among either borrowers or lenders is necessary for a collapse to take place. Finally, the direction of causation, in terms of propagators of the crisis, appears to run through credit channels by eroding depository institution real equity values. Only if institutions fail is the money supply affected.

In the **random** shock model of Diamond and Dybvig (1983), the crisis does not result from asset mispricing or from rational economic behavior but rather from an exogenous event. Since there is only one bank in the economy in this model, runs take the form of flights to currency (or more precisely, the consumption good) and not to other healthy banks. The panic is due solely to the existence of the sequential servicing requirement discussed above and the fact that bank assets are not perfectly liquefiable.

Like the micro **random** shock models, the asymmetric information models do not rest upon systematic ex ante asset mispricing or other problems of bank behavior. Changes in expectations and market assessment of bank asset quality, combined with the opaqueness of bank balance sheets and sequential servicing, make runs a rational customer response.

Empirical Evidence on Systemic Risk

When examined carefully, many of these alternative explanations of panics and **financial** crises appear to overlap, differing only slightly in their details. Separating them empirically can therefore be very difficult. Empirical tests of various hypotheses about **financial** crises and panics have generally focused on the National Banking Era and the period of the Great Depression. The reason for studying these periods is that no broad-based panics have occurred since (in part because of the existence of federal deposit insurance and lender-of- last-resort actions followed by the Federal Reserve). In this section, the empirical evidence is examined to determine which of the models appear to be more consistent with the data.

The question of whether this empirical work provides useful insights or is relevant today is a legitimate one, given the changes in **financial** structure and markets, the rise of technology, the proliferation of information, and the globalization of markets. This issue will be addressed in the next section.

The **Random** Shock and **Financial** Fragility Hypothesis. Given the lack of precision in specifying the models, does the evidence suggest that one or more of the models may be correct? With respect to the macro models, critics of the Minsky **financial** fragility hypothesis argue that it does not yield testable hypotheses and is inconsistent with the data (see Sinai 1977; Lintner 1977; Mishkin 1991; Schwartz 1986). As mentioned previously, for the hypothesis to hold, a sequence of several factors must be present: debt burdens increasing faster than the growth of income or capital stock, interest rates and nominal asset prices increasing rapidly, debt maturities at depository institutions becoming shorter, an exogenous shock occurring to cause a change in expectations, and, finally, governments failing to intervene in ways that would provide a soft landing to any asset revaluation that must accompany the change in expectations.

Unfortunately, data do not readily exist for examining a number of the conditions Minsky sets forward. As an alternative Table 1 lists the periods of economic recession with information on when panics took place and, where possible, what possible shocks may have existed. Looking first at the timing of the panics relative to the peaks and troughs of the business cycles shows that in only one instance was there a panic before the peak of the business cycle. In most cases, the panic occurred anywhere from three to six months after the business cycle had peaked. Such long lags would seem to be logically inconsistent with Minsky's view.

Mishkin (1991) devotes considerable attention to the rate pattern and

to risk premiums and their relationship to the onset of panics. In general, the spread between rates on high- and low-quality bonds rose before the panic began. However, these spreads generally widened after the recession started rather than before as the Minsky hypothesis would require.13

The Asymmetric Information Hypothesis versus Micro **Random** Shock Models. Gorton (1988), Mishkin (1991), and Donaldson (1992) specifically investigate the information asymmetric hypothesis in detail. Examining the National Banking Era and the post-Federal Reserve Era through 1933, Gorton models depositor behavior in terms of the currency/ deposit ratio. He poses two questions. First, if panics are **random** events, then is the model predicting a different currency/deposit ratio during panic periods than exists in other times? Second, are panics predictable in terms of movements of perceived risk? From these two questions Gorton suggests the following testable hypothesis: "Movements in variables predicting deposit riskiness cause panics just as such movements would be used to price such risk at all other times. This hypothesis links panics to occurrences of a threshold **value** of some variable predicting the riskiness of bank deposits" (1988, 751). Such predictive variables might be extreme seasonal fluctuations, unexpected failure of a large corporation (usually a **financial** corporation), or a major recession.

(Table Omitted)

Captioned as: TABLE 1

(Table Omitted)

Captioned as: PANIC DATES AND CUASES

(Table Omitted)

Captioned as: PANIC DATES AND CAUSES

A third question Gorton asks is whether certain predictors of risk stand out as important predictors of panics. Finding no evidence that panics are **random** events, he concludes that there is strong support for the asymmetric information hypothesis. Furthermore, panics appear to be predictable ex ante. Evidence also suggests that recession, and not a triggering bank failure, is the critical factor in determining whether a panic will occur. Gorton explains: "the recession hypothesis best explains what prior information is used by agents in forming conditional expectations. Banks hold claims on firms and when firms begin to fail (a leading indicator of recession), depositors will reassess the riskiness of deposits" (1988, 778). In short, causation seems directed from the real sector to the **financial** sector rather than vice versa.

Donaldson (1992) extends Gorton's **analysis** using a somewhat different specification of the model and weekly data between 1867 and 1907 to determine whether panics are systematic and predictable events. Unlike Gorton, Donaldson rejects the conclusion that panics are systematic events and argues that the data are more consistent with the **random** shock model than the asymmetric information model.14 However, for the panics of 1914 and 1933 (which required expansion of the money supply during crisis periods), he finds behavioral patterns of earlier panics had been dampened. Given that the later panics followed the creation of the Federal Reserve in 1913 and passage of the AldrichVreeland Act of 1912, this finding suggests that government involvement to increase liquidity can truncate panic situations. He concludes that panics are therefore special events. But he also finds evidence that panics are more likely to occur when seasonal and cyclical factors are present.

Mishkin (1991) formulates the asymmetric information hypothesis somewhat differently. He argues that key variables help to capture differences in depositor assessment of bank risk. In particular, during periods of **financial** distress high-quality firms will be less affected and lenders will have less uncertainty about the riskiness of such firms than they will have for low-quality firms. To the extent that these risks are priced, an important index of asymmetric information uncertainty should be captured by the spread between the rates on high- and low-quality bonds, by stock prices (as a measure of net worth and collateral

value), and by interest rates (as a measure of agency costs and adverse selection). His **analysis**, like that of Gorton (1988), supports the information asymmetries hypothesis to the extent that the proxy variables are in fact good proxies. He concludes that most **financial** crisis periods begin with an increase in interest rates and a widening of the spread between high- and low-quality bonds and a decline in stock prices, rather than with a panic. "Furthermore," Mishkin observes, "a **financial** panic was frequently immediately preceded by a major failure of a **financial** firm, which increased uncertainty in the marketplace" (1991, 97). He also asserts that the information hypothesis offers a better explanation than the macro theories of **financial** fragility for the pattern of rate spreads and stock market movements both before and after a panic as well as the panic's likely occurrence.

Finally, Park (1991) argues that the provision of bank-specific information overcame the information asymmetries that played a role in runs on banks. In particular, by analyzing the panics of 1873, 1884, 1893, 1907, and 1933 he concludes that clearinghouse and government intervention were effective devices in settling panics but only when they provided information on bank-specific solvency.15 In the panic of 1884, a run was abated following certification of solvency by the Comptroller of the Currency and by subsequent extensions of clearinghouse certificates to Metropolitan National Bank, which was the bank suffering the greatest withdrawals.

The panic of 1893 followed a long period of depression during which banks suffered prolonged periods of withdrawals of gold and uncertainty about U.S. adherence to the gold standard. Gold hoarding culminated in suspension of convertibility, and repeal of the Sherman Silver Act was promised by the president. Banks lifted the suspension of convertibility, and the runs stopped. Because no systematic attempt was made to release information on individual banks, public confidence in all banks remained low until the source of uncertainty-lack of confidence in U.S. maintenance of the gold standard-was removed. Park (1991) interprets the Comptroller of the Currency's certification of individual bank solvency before their reopening following the panic of 1893 as the major information factor that quelled depositor uncertainty.

In the panic of 1907, the problem began with runs on individual New York banks and trust companies that had been directly or indirectly associated with a failed attempt to corner the market in copper stocks. Only intervention by the New York Clearing House Association, which attested to the solvency of banks experiencing runs and provided **financial** assistance, resolved the situation. Again, release of firm-specific information appeared to have addressed the information asymmetries and helped stabilize the crisis." Unlike other cases, in the panic of 1907 runs did not affect all banks, and, indeed, some New York Clearing House member banks experienced reserve inflows (Park 1991).

Transmission Models. Neither the basic **random** shock models nor the information asymmetry models specifically address the issue of which mechanisms transmit panics or **financial** crises through the economy. In fact, no models admit more than one institution, a condition that would be necessary to model customers simply transferring funds from an unhealthy to a healthy institution as distinct from retreating to currency.' The models provide no information on what, if any, real impacts such funds transfers among banks have. Nor have the models addressed when depositors will run on one bank and when they will run on the entire system.

Researchers have addressed the question of transmission mechanism more indirectly by attempting to generalize from the basic models. For example, Calomiris and Gorton (1991) maintain that it is the sequential servicing constraint imposed in the Diamond and Dybvig-type models that can induce banks to run on other banks. Such runs are especially likely when banks are geographically dispersed but are permitted to count interbank deposits as

legal reserves, as under the National Banking system. Two other regulatory constraints-restrictions on branching and on the payment of interest on interbank deposits-have also been regarded as important.18

The structure of the National Banking system prior to creation of the Federal Reserve in 1913 added a further source of instability to the economy. Under that system, legal reserves for National Banks included not only cash in vault but also deposits in Reserve City and Central Reserve City banks. In such a fractionalreserve banking system that has pyramiding of reserves, a run on an individual bank could more easily have systemic, systemwide effects. Shocks originating in the countryside, for example, could induce country banks to improve their liquidity positions by recalling interbank deposits from the Reserve City and Central Reserve City banks. Hence, panics were also endemic to the structure of the system as a whole, and it is clear how a panic or run in a rural region could blossom into a systemic crisis for healthy banks in Reserve Cities and Central Reserve Cities. Chari (1989) addressed this issue directly in considering a model of spatially separated banks. He argued that the most likely source of a shock that would cause country banks to withdraw reserve funds was seasonally related, with differences in currency demands rising significantly during planting and harvest times.

Calomiris and Gorton (1991) attempted to determine specifically whether panics were transmitted from rural areas through the National Banking system, as the **analysis** suggested, and also whether the patterns were more consistent with the **random** shock or information asymmetries models. They found that three important differences between the models have empirical implications.

The first concerns the origin of problems. The **random** shock model suggests that shocks would occur in rural areas because of seasonal demands for currency. In contrast, the asymmetric information model implies that adverse economic news related to asset-quality problems would precede a panic.

Second, the two theories would seem to predict different patterns of failures during a crisis. The asymmetric information model suggests that banks whose asset portfolios were closely linked to the specific shock would be more prone to failure whereas the **random** shock model would predict that failures would be experienced in the areas suffering currency withdrawals.

Finally, the models differ in the conditions required to resolve a crisis. In the **random** shock model, the key to resolving a panic is liquefication of assets. In the asymmetric information model, it is the effectiveness of mechanisms initiated to resolve depositor uncertainty about bank solvency.

Calomiris and Gorton's exhaustive investigation of the sources of panics between 1873 and 1907 led them to reject the idea that seasonal money-demand shocks were the cause of banking panics. Rather, their **analysis** suggests that panics originated in bad economic news and bank vulnerability to that news. Moreover, their inspection of failure patterns shows virtually no support for the **random** shock model. Finally, they conclude that in terms of resolving crises, the mere availability of currency, which would provide the ability to liquefy assets, was not sufficient to stop panics during the periods studied. Again, this conclusion suggests that the asymmetric information model was more consistent with the data than was the **random** shock model.

Smith (1991) provides some specific evidence on country banks' behavior vis-a-vis their holdings of cash reserves as compared with reserves held in the form of interbank deposits when panics occurred. He provides **analysis** of some anecdotal and other evidence, derived mostly from Sprague ([1910] 1968), about the behavior of Reserve City banks during the crises of 1873, 1893, 1907, and 1930-33.

Smith describes the situation leading up to the panic of 1873, indicating that interbank deposits were concentrated in seven of the New York City banks. These interbank deposits constituted about 45 percent of

the sources of funds for the New York banks and were the base upon which their bond holdings and loans were built. These banks were clearly vulnerable to demands by country banks for withdrawal of reserves, and funds were especially tight in the few months before the crisis. When the key triggering events occurred (see Table 1), a combination of circumstances made the crisis severe. In addition to having virtually no excess reserves, several of the banks were in weak **financial** condition. As subsequent events would prove, several had been the victims of fraud and defalcations, probably accounting in part for their **financial** weakness. Clearly, however, the institutions' problems stemmed primarily from reserve withdrawals and their inability to call in loans in that economic environment rather than from major credit problems in the New York Central Reserve City banks.

The Reserve City banks experienced similar problems caused by currency outflows during the panics of 1893 and 1907. Thus, it seems clear that reserve outflows, coupled with the lack of excess currency reserves at the Central Reserve City banks in New York and Chicago, forced contractions in loans and finally resulted in the suspension of currency payments. Smith notes that currency suspension was the prime transmission mechanism of panics once a triggering mechanism occurred. He also concludes that the problems during the 1930-33 period originated in the rural agricultural areas as well and were intimately intertwined with the correspondent banking system.

Despite a fairly clear pattern in the transmission mechanism of panics emanating from large reservedeposit withdrawals (rather than from uncertainties about credit quality in Reserve City banks, as the Minsky hypothesis would imply), a number of questions remain. For example, Tallman (1988) indicates that looking at the data over longer time periods does not suggest a clear linkage between the incidence of panics and either increases in currency demand relative to deposits or contractions in loans. He presents evidence that loan contractions occurred at several **intervals** during the period between 1893 and 1907, for example, that exceeded the declines during periods when panics occurred. Similarly, during some periods of time between 1873 and 1930 the number of bank failures far exceeded those observed during panic periods. Finally, Tallman provides aggregate data on the growth in loans relative to high-powered money and on the growth of manufacturing output between 1873 and 1914. Two observations are important. First, loans do increase in the years prior to panic periods, but the panics occur after loan growth has fallen significantly. Second, numerous periods during the **interval** show the same patterns in loan and output growth and decline but are not accompanied by a panic. These aggregate data do not reveal whether there are differences in the loancontraction periods in terms of their concentration in particular parts of the country during episodes of panic and nonpanic periods.

Causal Direction. The research evidence seems to indicate fairly consistently that the dynamics between **financial** panics and changes in real economic output begin in the real sector and move to the **financial** sector rather than starting in the **financial** sector. There are no examples in U.S. history of the economy operating at high levels of output when a **financial** crisis occurred that resulted in a contraction in the real economy. As the discussion in the previous sections suggests, however, banks were sometimes under pressure and were forced to call in loans. It seems reasonable to assume that once problems in the **financial** sector become severe, there could be negative feedback effects to the real sector.

Indeed, Bernanke (1983) has made precisely this point. **Financial** crises can have real effects outside the normal reserve/loan transmission mechanism because of the disruptions to the intermediation process. Bank failures disrupt borrower/lender relationships and make attaining **financing** more difficult and costly. But this observation should not obscure the fact that **financial** crises

are better viewed as creatures of recession and economic downturns rather than primary causal agents precipitating the downturns.

The Role of Government. A substantial body of evidence indicates that government actions have played significant roles in contributing to crises as well as in mitigating them. For example, Sprague ([1910] 1968) notes that lack of access to a reliable lender of last resort to provide short-term liquidity can help escalate a period of **financial** tightness into one of crisis. Friedman and Schwartz (1963) argue that several Federal Reserve actions during the Great Depression contributed to both its duration and magnitude. For instance, they observe that the Federal Reserve's failure to liquefy the assets of many small nonmember institutions (the Fed was not obligated to lend to nonmember banks), together with its insistence that it would lend only upon sound collateral, added to the number of bank failures. This policy, in conjunction with the Fed's attempt to adhere to the rules of the gold standard, contributed to a 33 percent decline in the money supply and clearly exacerbated the severity of the recession.

While it has become fashionable to criticize the Fed for its policy failures during the Great Depression, it is also the case that government interference affected **financial** soundness long before the Fed was created. For example, during the National Banking Era the monetary base was tied, except for a period of suspension, to gold and silver specie monies through the Treasury.19 When the United States adhered to the gold standard, fluctuations in the gold supply expanded and contracted the monetary base, directly affecting banks' lending behavior. Decisions about how much in the way of international gold flows would be permitted before conversion could be suspended was a matter of Treasury and government policy. European central banks, and to a lesser extent the U.S. Treasury, often intervened to prevent loss of gold reserves by raising short-term interest rates. Government policies frequently exacerbated gold flows and, by implication, induced fluctuations in the monetary base. For example, following passage of the Sherman Silver Purchase Act in 1890, foreigners' concern that the United States would remain on the standard precipitated gold outflows and contributed to the panic of 1890.20

Tallman and Moen note that each panic after 1897 was preceded by unusual gold flows. They conclude that political uncertainties concerning the U.S. commitment to the gold standard were important influences on gold flows and, hence, the U.S. monetary base. Political conditions outside the United States also affected gold flows. For example, in 1907 the Bank of England responded to problems in the London money markets by raising its discount rate to stem potential speculative outflows of gold to the United States. At that time London was the most important market for discounting U.S. trade bills. The increase in the discount rate not only disrupted the flow of gold to the United States but also discouraged the discounting of trade bills and caused a liquidity crisis in the United States.

Since the debacle of the Great Depression, U.S. intervention in markets has often had as its objective providing liquidity to avoid a crisis. Numerous examples exist of emergency liquidity having been provided through the efforts of the Federal Reserve either directly or indirectly, such as during the Penn Central scare, the Chrysler problem, the collapse of Drexel-Burnham, and the failure of Continental Illinois Bank, to name just a few. The Federal Reserve has on occasion attempted to provide liquidity not only to cushion problems in interbank markets but also to prevent disruptions in other markets.

Relevance in Today's World

It can be argued that inferences about **financial** crises and systemic risk drawn from study of the banking situation during the National Banking Era and early 1900s are not particularly meaningful or relevant in today's economic environment. Pyramiding of legal reserves in private banks is not a structural feature of the present reserve requirement regime. Markets are no longer isolated, and informationavailability problems that might have resulted in the past in information asymmetries have been

reduced significantly. Communications technology and new instruments have increased the liquidity of all banking assets and have given rise to new markets that make the kinds of liquidity crises that occurred in the National Banking Era unlikely today. Furthermore, the United States has abandoned the gold standard, and thus the domestic money supply is not subject to the **random** fluctuations and shocks that it was vulnerable to under strict adherence to the gold standard rules. Deposit rate ceilings of the 1930s have been phased out, and branching restrictions, which essentially prevented institutions from achieving geographical diversification, are a thing of the past.

Certainly the focus on protecting the money supply from sudden shocks is no longer of prime policy concern for three reasons. First, it seems unlikely that significant runs to currency will occur (see Kaufman 1988). Deposits still have large advantages over currency for a variety of purposes, and there are many banks to chose from. Runs on individual banks would simply transfer reserves from one institution to another. Second, Federal Reserve policy is likely to provide emergency liquidity to prevent such runs from disrupting other institutions. Finally, while still accounting for the bulk of payment items, checks and currency are no longer the dominant forms in terms of dollar volume of transactions in the economy. The concerns and risks have shifted to other sectors of the payments system that did not exist during the National Banking Era.

Today, the payments system is larger, has many more components (both private and public), and is subject to different risks than in the past. The check/ demand deposit system, which accounts for the bulk of individual payments (except for currency), and the one that the present regulatory structure was primarily designed to protect, is small in terms of the dollar volume of payments. The rest are made in the form of computerized transfers of reserve balances on the Federal Reserve's Fedwire system and the privately owned Clearing House Interbank Payments System (CHIPS) and in the form of automated clearinghouse (ACH) transactions. Payments on the former two systems account for about 85 percent of the dollar **value** of transactions. Closely related to these systems are the automated transfers of book-entry Treasury securities, which also take place on Fedwire and involve substantial volumes of transactions.

Finally, as markets have become increasingly global, timing differences and differences in clearing and settlement conventions can add temporal and other dimensions to credit risks not always found in the domestic markets that characterized earlier times. Many other significant sources of uncertainty can also be identified in the clearing and settlement processes in modern **financial** markets (see, for example, Eisenbeis 1997 and McAndrews 1997).

Maintaining the integrity of payment flows is a substantially more complicated and difficult problem today than protecting the stock of demand deposits for a number of reasons. First, given the large size of transactions in the system and the size of the system itself, the resources required to support unwinding even a shortrun problem may be enormous and could exceed the capacity of private participants to self-insure. Second, because the transactions are electronic and occur instantaneously, monitoring them and the net position of each participant is critical to controlling participants' credit risk exposure. Third, when the international activities of U.S. banks and the links between the U.S. domestic payments system and foreign banking organizations are recognized, it becomes difficult to conceive of ensuring domestic **financial** stability without also ensuring international **financial** stability.

Clearly, different types of uncertainties exist with respect to systemic risk exposures today than existed in the past. There is also reason to believe that liquidity problems for borrowers may be significantly different than they were for borrowers in the 1800s. The growth of new mortgage lending instruments and, particularly, the development of home equity lines of credit provide ways for borrowers

instantaneously to liquefy previously illiquid assets during tight times. While this ability to liquefy assets more easily may enable borrowers to maintain payments on outstanding debts and lessen the severity of the credit component of a recession, it also suggests introduction of a new discontinuity that might systematically transfer risks to the banking system at a critical trigger point. If during times of **financial** distress borrowers draw down lines on home equity and similar lines of credit and are then forced into default, the burdens of these defaults will be shifted to the providers of the home equity lines. Should these losses be large, capital might be wiped out, with few options available to lenders to avoid the costs of those defaults. Examples of similar impacts in commercial and real estate lending markets occurred when commercial paper borrowers drew down banks' back-up commitments during the Penn Central and Real Estate Investment Trust (REIT) crises.

Summary and Conclusions

This article has investigated the various theories of **financial** panics and crises with particular emphasis on the links between credit and deposits. The survey suggests that panics are not **random** events, as some of the theories may suggest, but neither are they perfectly predictable. Nevertheless, it does appear that information asymmetries about the ability to liquefy deposits were a major contributing factor to banking panics in the past. Moreover, while panics do appear to be associated with recessions and deflationary periods, the direction of causation seems to run from the real sector to the **financial** sector rather than the other way around. It is not that **financial** crises cannot exacerbate economic declines; rather, they are not primary causal agents of recessions.

The **analysis** also suggests that government policies can affect the likelihood of a **financial** crisis as well as play a role in its solution. These considerations are as relevant today as they have been **historically**. At the same time, the article raises a cautionary note that the dynamics of crises and how they might play out may be significantly different in the future given recent, rapidly developing changes in the U.S. and world **financial** system.

Footnote:

1. For representative examples see Smith (1991), Kaufman (1995), Donaldson (1992), Bartholomew, Moe, andWhalen (1995), and Eichengreen and Portes (1991). See Benston and Kaufman (1995)for a review of the evidence on fragility. 2. Eichengreen and Portes (1991) require declines in real output for a true **financial** crisis to occur 3. Before 1910, however, the most common yield curve in the United States was downward-sloping. 4. It is generally argued that the theory as put forth bg Minsky is not a unified theory that yields testable hypotheses. See, for example, Sinai (1977), Lintner (1977), Mishkin (1991), and Schwartz (1986).

Footnote:

5. Minsky argues that the ability to intervene is directly correlated with the size of government; and big government, with its revenue capacity, has the resources to support, through fiscal and monetary policies, a longer run-up of leverage. Also, through its lender-of-last-resort capabilities, it can soften the landing during an exogenous shock period by supporting a gradual rather than precipitous liquidation of assets. It thereby avoids the corresponding collapse of credit, bank failures, and destruction of the money supply.

6. For other examples of models in this mode see Haub rich and King (1984), Cone (1983), Jacklin (1987), Wallace (1988), Bhattacharya and Gale (1987), Smith (1991), and Chari (1989).

7. In the Diamond and Dybvig (1983) model there is really no nonbank money in circulation. Individuals deposit a real consumption good in the bank in exchange for a deposit or warehouse receipt. This consumption good is close, but not identical, to specie.

In early US. banking, it was not un.common for notes issued by out-of-area banks to trade at discounts, which reflected several factors,

including transportation and transaction costs, lack of information on the issuing bank, and uncertainties about the creditworthiness of the issuing bank. This lack of par clearance in no way affected the ability of state bank notes to function as money.

8. For discussions of the evidence on runs see Kaufman (1988) and Gorton (1987). 9. Because of the way the model is constructed, runs necessarily have an adverse impact on the real economy. 10. For a discussion of these early bank runs see Kaufman (1988) or Bryant (1980). 11. Clearinghouses and other banks in the region often provided temporary credit to institutions experiencing liquidity problems (see Kaufman 1988). Kaufman (1994) notes that bank capital ratios were substantially higher during this period than they were after deposit insurance was introduced.

12. Although Schwartz articulates this view, she clearly does not believe it is correct or that the policies designed to protect against the events are appropriate.

Footnote:

13. The exception is the panic of 1873.

14. As a robustness test, he also reruns the **analysis** using monthly data as Gorton does and gets similar results to those found by Gorton. He concludes that monthly data are too spaced out to provide a sharp test of the hypothesis. 15. A more complete test of the Gorton-Mishkin-Park hypothesis about information asymmetries would be provided by examining fund,flows from individual solvent and insolvent institutions. Relying upon aggregate statistics can be only circumstantial, not conclusive.

16. The Roosevelt administration, following the declaration of the bank holiday on March 6, 1933, employed this same policy. 17 Smith (1991) does provide a model in which banks are permitted to hold funds at a Reserve City bank. Bhattacharya and Gale (1987) provide a model with geographically dispersed depositors and banks. Again, however these models only look at the interdependence among banks through the interbank deposit markets.

18. See also the discussions in Haubrich (1990), Bordo (1986), and Williamson (1989). All emphasize the advantages over US. banks that banks in Canada and other countries that permitted branching had in weathering panics. Calomiris and Schweikart (1991) have explored in detail for the United States the effects that structure had on failure rates in different states with different branching statutes. They show that branch banks had both lower failure rates and in general paid lower premiums on their notes during the crisis of 1857 than banks in other parts of the country.

Footnote:

19.Specifically, the monetary base included gold coin, gold certificates backed 100 percent by gold, silver dollars, silver certificates, other small silver coins, US. notes and other Treasury fiat, and national bank notes. See Tallman and Moen (1993). 20. Tallman and Moe (1993) indicate that this uncertainty was greatly reduced with the discovery of large gold supplies in the late 1890s.

Reference:

REFERENCES

Reference:

AKERLOF, GEORGE. 1979. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism. Quarterly Journal of Economics 84:488-500.

BARTHOLOMEW, PHILIP F., LARRY R. MOE, AND GARY WHALEN. 1995. "The Definition of Systemic Risk." Office of the Comptroller of the Currency. Presented at the seventieth annual Western Economic Association International Conference, San Diego, California, July.

BENSTON, GEORGE J., AND GEORGE G. KAUFMAN. 1995. "Is the Banking and Payments System Fragile?" Journal of **Financial** Services Research 9 (December): 209-40.

BERNANKE, BEN S. 1983. "Non-Monetary Effects of the **Financial** Crisis in the Propagation of the Great Depression." American Economic Review 73 (June): 14-31.

BERNANKE, BEN, AND HAROLD JAMES. 1991. "The Gold Standard, Deflation, and **Financial** Crisis in the Great Depression: An International Comparison." In **Financial** Markets and **Financial** Crises, edited by R. Glenn Hubbard, 3368. Chicago: University of Chicago Press.

Reference:

BHATTACHARYA, SUDIPTO, AND DOUGLAS GALE. 1987. "Preference Shocks, Liquidity, and Central Bank Policy." In New Approaches to Monetary Economics, edited by William Barnett and Kenneth Singleton, 69-88. New York: Cambridge University Press.

BORDO, MICHAEL. 1986. "**Financial** Crises, Banking Crises, Stock Market Crashes, and the Money Supply: Some International Evidence" In **Financial** Crises and the World Banking System, edited by Forest Capie and Geoffrey E. Wood, 190-248. London: Macmillan.

BRYANT, JoHN.1980. "A Model of Reserves, Bank Runs, and Deposit Insurance." Journal of Banking and **Finance** 4:335-44.

Reference:

CALOMIRIS, CHARLES W., AND GARY GORTON. 1991. The Origins of Banking Panics: Models, Facts, and Bank Regulation." In **Financial** Markets and **Financial** Crises, edited by R. Glenn Hubbard, 109-73. Chicago: University of Chicago Press.

CALOMIRIS, CHARLES W., AND LARRY SCHWEIKART. 1991. The Panic of 1857: Origins, Transmission, and Containment." Journal of Economic History 51:807-34.

CHARI, VV. 1989. 'Banking without Deposit Insurance or Bank Panics: Lessons from a Model of the U.S. National Banking System." Federal Reserve Bank of Minneapolis Quarterly Review 13 (Summer): 3-19.

CHARI, V.V., AND RAVI JAGANNATHAN. 1984. "Banking Panics, Information, and Rational Expectations Equilibrium." Northwestern University. Photocopy.

CONE, KENNETH. 1983. Regulation of Depository Institutions." Ph.D. diss., Stanford University.

DIAMOND, DOUGLAS, AND PHILIP DYBVIG. 1983. "Bank Runs, Liquidity, and Deposit Insurance." Journal of Political Economy 91 (June): 401-19.

Reference:

DONALDSON, R. GLEN. 1992. "Sources of Panics: Evidence from Weekly Data." Journal of Monetary Economics 30 (November): 277-305.

EICHENGREEN, BARRY, AND RICHARD PORTES. 1991. The Anatomy of **Financial** Crises." In **Financial** Markets and **Financial** Markets and **Financial** Crises, edited by R. Glenn Hubbard, 10-58. Chicago: University of Chicago Press.

EISENBEIS, ROBERT A. 1997. "International Settlements: A New Source of Systemic Risk?" Federal Reserve Bank of Atlanta Economic Review 82 (Second Quarter): 44-50.

FRIEDMAN, MILTON, AND ANNA J. SCHWARTZ. 1963. A Monetary History of the United States, 1867-1960. Princeton, N.J.: Princeton University Press.

GORTON, GARY. 1987. "Banking Panics and Business Cycles: Data Sources, Data Construction, and Further Results." The Wharton School, University of Pennsylvania. Photocopy.

_____.1988. "Banking Panics and Business Cycles. Oxford Economic Papers 40:751-81.

HAUBRICH, JOSEPH. 1990. Non-Monetary Effects of **Financial** Crises: Lessons from the Great Depression in Canada." Journal of Monetary Economics 25 (March): 223-52.

HAUBRICH, JOSEPH, AND ROBERT KING. 1984. "Banking and Insurance." NBER Research Working Paper No. 1312.

JACKLIN, CHARLES. 1987. "Demand Deposits, Trading Restrictions, and Risk Sharing." In Contractual Arrangements for Intertemporal Trade, edited by Edward D. Prescott and Neil Wallace, 26-47. Minneapolis: University of Minnesota Press.

KAUFMAN, GEORGE G. 1988. "Bank Runs: Causes, Benefits, and Costs." Cato Journal 7 (Winter): 559-87.

_____. 1994. "Bank Contagion: A Review of the Theory and Evidence.'

Journal of **Financial** Services Research 8:123-50.

_____. 1995. "Comment on Systemic Risk." In Research in **Financial** Services, vol. 7, edited by George G. Kaufman. Greenwich, Conn.: JAI Press.

KINDLEBERGER, CHARLES P. 1978. Man.ias, Panics, and Crashes. New York: Basic Books.

LINTNER, JOHN. 1977. "Discussion." In **Financial** Crises: Institutions and Markets in a Fragile Environment, edited by Edward I. Altman and Arnold W. Sametz, 204-207. New York: John Wiley and Sons.

McANDREws, JAMES J. 1997. "Banking and Payment System Stability in an Electronic Money World." Federal Reserve Bank of Philadelphia Working Paper No. 97-9.

MINSKY;, HYMAN P. 1982. "The **Financial**-Instability Hypothesis: Capitalist Processes and the Behavior of the Economy. In **Financial** Crises: Theory, History, and Policy, edited by Charles P. Kindleberger and Jean-Pierre Laffargue, 13-38. Cambridge: Cambridge University Press.

Reference:

MISHKIN, FREDERIC S. 1991. "Asymmetric Information and **Financial** Crises: A **Historical** Perspective." In **Financial** Markets and **Financial** Crises, edited by R. Glenn Hubbard, 69-108. Chicago: University of Chicago Press.

PARK, SANGKYN. 1991. Bank Failure Contagion in **Historical** Perspective." Journal of Monetary Economics 28 (October): 271-86.

SCHWARTZ, ANNA. 1986. "Real and Pseudo **Financial** Crises." In **Financial** Crises and the World Banking System, edited by Forest Capie and Geoffrey E. Wood, 11-31. London: Macmillan.

SINAI, ALLEN. 1977. Discussion." In **Financial** Crises: Institutions and Markets in a Fragile Environment, edited by Edward I. Altman and Arnold W. Sametz, 187-203. New York: John Wiley and Sons.

SMITH, BRUCE D. 1991. Bank Panics, Suspensions, and Geography: Some Notes on the 'Contagion of Fear' in Banking." Economic Inquiry 29 (April): 230-48.

Reference:

SpRAGUE, O.M.W. [1910] 1968. History of Crises under the National Banking System. National Monetary Commission. Reprint. New York: Augustus M. Kelley.

TALLMAN, ELLIS W. 1988. "Some Unanswered Questions about Bank Panics." Federal Reserve Bank of Atlanta Economic Review 73 (November/December): 2-21.

TALLMAN, ELLIS W., AND JON MOEN. 1993. "Liquidity Shocks an **Financial** Crises during the National Banking Era." Federal Reserve Bank of Atlanta Working Paper 93-10, August.

WALLACE, NEIL. 1988. "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously." Federal Reserve Bank of Minneapolis Quarterly Review 12 (Fall): 3-16.

WILLIAMSON, STEVEN. 1989. "Bank Failures, **Financial** Restrictions, and Aggregate Fluctuations: Canada and the United States, 1870-1913." Federal Reserve Bank of Minneapolis Quarterly Review 13 (Summer): 20-40.

Special Features: Charts; References
Classification: 9190 (CN=United States); 1130 (CN=Economic theory); 8100 (CN=Financial services industry); 9130 (CN=Experimental/Theoretical)
Descriptors: Risk; Economic models; Economic theory; Bank failures; Bank deposits; Studies; Economic

impact; Credit
Geographic Names: US

36/9/9 (Item 9 from file: 268)
00315937
Credit, equity, and mortgage refinancings
Peristiani, Stavros; Bennett, Paul; Monsen, Gordon; Peach, Richard; Raiff, Jonathan
Economic Policy Review , v 3 , n 2 , p 83-99 , Jul 1997 **Document Type:** Journal Article **Journal Code:** EPV **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 08404

Abstract:

Studies examining refinancing behavior are finding more evidence that differences in homeowners' ability to qualify for new mortgage credit, and differences in the cost of that credit, account for a significant part of the variation in that behavior. A study introduces quantitative measures of individual homeowner credit histories to the loan-level **analysis** of the factors influencing the probability that a homeowner will refinance. In addition to credit histories, changes in individual homeowner's equity and in the overall lending environment are analyzed. The findings show that, other things being equal, the worse a homeowner's credit rating, the lower the probability of refinancing. Evidence is provided of a change in the lending environment that, all else being equal, has increased the probability that a homeowner will refinance. These findings are important from an investment risk management perspective because they confirm that the responsiveness of mortgage cash flows to changes in interest rates will also be significantly influenced by the credit and equity conditions of individual borrowers.

Text:

Homeowners typically have the option to prepay all or part of the outstanding balance of their mortgage loan at any time, usually X s without penalty. However, unless homeowners have sufficient wealth to pay off the balance, they must obtain a new loan in order to exercise this option. Studies examining refinancing behavior are finding more and more evidence that differences in homeowners' ability to qualify for new mortgage credit, as well as differences in the cost of that credit, account for a significant part of the observed variation in that behavior. Therefore, individual homeowner and property characteristics, such as personal credit ratings and changes in home equity, must be considered systematically, along with changes in mortgage interest rates, in the **analysis** and prediction of mortgage prepayments.

Early research into the factors influencing prepayments focused almost exclusively on the difference between the interest rate on a homeowner's existing mortgage and the rates available on new loans. This approach arose in part because researchers most often had to rely on aggregate data on the pools of mortgages serving as the underlying collateral for mortgage-backed securities (for example, see Schorin 1992). More recent research, however, has broadened the scope of this investigation through the utilization of

loan-level data sets that include individual property, loan, and borrower characteristics.

This article significantly advances the literature on mortgage prepayments by introducing quantitative measures of individual homeowner credit histories to the loan-level **analysis** of the factors influencing the probability that a homeowner will refinance. In addition to credit histories, we include in the **analysis** changes in individual homeowner's equity and in the overall lending environment. Our findings strongly support the hypothesis that, other things being equal, the worse a homeowner's credit rating, the lower the probability that he or she will refinance. We also confirm the finding of other researchers that changes in home equity strongly influence the probability of refinancing. Finally, we provide evidence of a change in the lending environment that, all else being equal, has increased the probability that a homeowner will refinance.

These findings are important from an investment risk management perspective because they confirm that the responsiveness of mortgage cash flows to changes in interest rates will also be significantly influenced by the credit and equity conditions of individual borrowers. Moreover, evidence overwhelmingly indicates that these conditions are subject to dramatic changes. For example, although the sharp rise in personal bankruptcies since the mid-1980s (Chart 1) partly reflects changes in laws and attitudes, it nonetheless suggests that credit histories for a growing segment of the population are deteriorating. Furthermore, home price movements, the key determinant of changes in homeowners' equity, have differed considerably over time and in various regions of the country. Indeed, in the early to mid-1990s home price appreciation for the United States as a whole slowed dramatically while home prices actually fell for sustained periods in a few regions (Chart 2).

In short, as mortgage rates fell during the first half of the 1990s, many households likely found it difficult, if not impossible, to refinance existing mortgages because of poor credit ratings or erosion of home equity.1 Consequently, the prepayment experience of otherwise similar pools of mortgage loans may vary greatly depending on the pools' proportions of credit- and/or equity-constrained borrowers.

Our findings also contribute to an understanding of how constraints on credit availability affect the transmission of monetary policy to the economy (for example, see Bernanke [1993). Fazzari, Hubbard, and Petersen (1988) and others have found that investment expenditures by credit-constrained businesses are especially closely tied to those firms' cash flows and are relatively insensitive to changes in interest rates, reflecting constraints on their ability to obtain credit. Analogously, we find credit- and/or equity-constrained homeowners to be less sensitive to changes in interest rates because of their limited access to new credit, thereby short-circuiting one channel through which lower interest rates improve household cash flows and stimulate the economy.

PREVIOUS LOAN-LEVEL RESEARCH ON MORTGAGE PREPAYMENTS

Recognition that individual loan, property, and borrower characteristics, in addition to changes in interest rates, play a key role in determining the likelihood of a mortgage prepayment has spawned a relatively new branch of research based on loan-level data sets. This research has generally focused on the three major underwriting criteria that mortgage lenders consider when deciding whether to extend credit: equity (collateral), income, and credit history.

(Graph Omitted)

Captioned as: Chart 1

(Graph Omitted)

Captioned as: Chart 2

However, past studies have only investigated the effects of changes in homeowners' equity and income on their ability to prepay. For example, Cunningham and Capone (1990)-using a sample of loans secured by properties in the Houston, Texas, area-estimated post-origination loan-to-

value (LTV) ratios and post-origination paymentto-income ratios based on changes in regional home prices and incomes.2 They concluded that post-origination equity was a key determinant of the termination experience of those loans (they found an inverse relationship for defaults and a positive relationship for refinancings and home sales), whereas post-origination income was insignificant. Caplin, Freeman, and Tracy (1993), using a sample of loans secured by properties in six states, also found evidence of the importance of home equity in influencing the likelihood of mortgage prepayment. They assessed the effect of post-origination equity by dividing their sample into states with stable or weak property markets (using transaction-based home price indexes for specific metropolitan statistical areas) and according to whether the loans had high or low original LTV ratios. Consistent with the hypothesis that changes in home equity play an important role in prepayments, the authors found that in states with weak property markets, prepayment activity was less responsive to declines in mortgage interest rates than in states with stable property markets.

In a related study, Archer, Ling, and McGill (1995) found that home equity had an important effect on the probability that a loan would be refinanced, and provided evidence that changes in borrower income are also a significant factor. The authors matched records from the 1985 and 1987 national samples of the American Housing Survey to derive a subsample of nonmoving owner-occupant households with fixed-rate primary mortgages, some of whom had refinanced, since the interest rate on their loan in 1987 was different from that reported in 1985. The authors' estimate of post-origination home equity was derived from the sum of the book **value** of a homeowner's entire mortgage debt, including second mortgages and home equity loans, divided by the owner's assessment of the current **value** of his or her property.3 In addition, a post-origination mortgage payment-to-income ratio, derived from the homeowner's recollection of total household income, was included as an explanatory variable. The authors found that, along with changes in interest rates, post-origination home equity and income were significant and of the expected sign.

This article goes beyond the existing literature in several important respects. Ours is the first study to investigate systematically the effect of the third underwriting criterion: homeowners' credit histories. Ours is also the first study to estimate post-origination equity by using county-level repeat sales home price indexes.4 These indexes are generally regarded as the best available indicator of movements in home prices over time. In addition, we employ a unique loan-level data set that not only provides information on credit history but also identifies the reason for prepayment: refinance, sale, or default (see box). The size of the data set allows very large samples to be drawn for major population centers as well as for the nation as a whole.

The data for this study were provided by the Mortgage Research Group (MRG) of Jersey City, New Jersey, which in the early 1990s entered into a strategic alliance with TRW-one of the country's three largest credit bureaus-to provide data for research on mortgage **finance** issues. Until late 1996, MRG maintained a data base, arranged into "tables," of roughly 42 million residential properties located in 396 counties in 36 states. The primary table is the transaction table, which is based on the TRW Redi Property Data data base. This table is organized by properties, with a detailed listing of the major characteristics of all transactions pertaining to each property. For the roughly 42 million properties covered, information is provided on 150 million to 200 million transactions. For example, if a property is purchased, a purchase code is entered along with key characteristics of the transaction, including date of closing, purchase price, original mortgage loan balance, and maturity and type of mortgage (such as fixed-rate, adjustable-rate, or balloon).

The characteristics of any subsequent transactions are also recorded, such as a refinancing of the original mortgage, another purchase of the

same property, and, for some counties, a default. The primary sources of this information are the records of county recorders and tax assessors, which are surveyed on a regular basis to keep the transaction data current.

A separate table contains periodic snapshots of the credit histories of the occupants of the properties. The data on credit histories are derived from TRW Information Services, the consumer credit information group of TRW. The data include summary measures of individuals' credit status as well as detailed delinquency information on numerous categories of credit sources. Individual records in the credit table can be linked to records in the transaction table on the basis of property identification numbers.

For our study, a sample from the larger data set was constructed in several stages: First, we selected groups of counties representing the 4 major regions of the country. In the East, we chose 4 counties surrounding New York City (Orange County in New York State, and Essex, Bergen, and Monmouth Counties in New Jersey). In the South, we chose 6 counties in central Florida (Citrus, Clay, Escambia, Hernando, Manatee, and Marion). In the Midwest, we chose Cook County and 5 surrounding counties in Illinois (Dekalb, DuPage, Kane, McHenry, and Ogle). In the West, we selected Los Angeles, Ventura, and Riverside Counties in California. Selecting these 4 diverse areas assured us that our statistical findings would be general rather than specific to a particular housing market. Furthermore, over the past decade, the behavior of home prices in the 4 regions has been quite different.

In the 19 counties examined, we identified for each property the most recent purchase transaction, going back as far as January 1984. The mortgages on some of these properties were subsequently refinanced, in some cases more than once, while other properties had no further transactions recorded through the end of our sample period, December 1994. (For multiple refinancings, we considered just the first one. In addition, we excluded from the sample loans that subsequently defaulted.) Thus, the sample consisted of loans that were refinanced and loans that were not refinanced as of the end of the sample period, establishing the zero/one, refinance/no-refinance dependent variable we then try to explain. (For refinanced loans, the new loan could be greater than, equal to, or less than the remaining balance on the old loan.) We limited our sample to fixed-rate mortgages outstanding for a year or more; the decision to refinance alternative mortgage types is more complex to model and is not treated in this study.

In the final step, MRG agreed to link credit records as of the second quarter of 1995 to a **random** sample of these properties. (Note that any information that would enable users of this data set to identify an individual or a property was masked by MRG.) The resulting sample consisted of 12,855 observations, of which slightly under one-third were refinanced.

Our sample is an extensive cross section, with each observation representing the experience of an individual mortgage loan over a well-defined time period. For example, assume that an individual purchased a house in January 1991 and subsequently refinanced in December 1993, an **interval** of 36 months. This window represents one observation or experiment in our sample. Our approach differs from that of most other studies on this topic in that the starting date, ending date, and time **interval** between refinancings are unique for each observation. Starting dates (purchases) range from January 1984 to December 1993, while time **intervals** (loan ages) range from 12 to 120 months. Therefore, our sample includes refinancings that occurred in the "refi wave" from 1986 to early 1987 as well as in the wave from 1993 to early 1994, although most are from the latter period. This diverse sample allows us to investigate whether the propensity to refinance has changed over time.

MODELING THE DECISION TO REFINANCE

When a homeowner refinances, he or she exercises the call option

imbedded in the standard residential mortgage contract. In theory, a borrower will exercise this option when it is "in the money," that is, when refinancing would reduce the current market **value** of his or her liabilities by an amount equal to or greater than the costs of carrying out the transaction. In fact, however, many borrowers with apparently in-the-money options fail to exercise them while others exercise options that apparently are not in the money. This heterogeneity of behavior appears to be due partly to differences in homeowners' ability to secure replacement **financing**. If an individual cannot qualify for a new mortgage, or can qualify only at an interest rate much higher than that available to the best credit risks, then refinancing may not be possible or worthwhile even though at first glance the option appears to be in the money.

While a decline in equity resulting from a drop in property **value** may rule out refinancing for some homeowners, refinancing may also not be possible or worthwhile because the homeowner's personal credit history is marginal or poor. This condition either prevents the borrower from obtaining replacement **financing** or raises the cost of that **financing** such that the present **value** of the benefits does not offset the transaction costs. Not only might the interest rate available exceed that offered to individuals with perfect credit ratings, but transaction costs might also be higher. In addition to paying higher out-of-pocket closing costs, the credit-impaired borrower may be asked to provide substantially more personal **financial** information and may face a substantially longer underwriting process.

Of course, other factors may explain this heterogeneity of refinancing behavior. For instance, homeowners often refinance when the option is not in the money in order to take equity out of the property. After all, mortgage debt is typically the lowest cost debt consumers can obtain, particularly on an after-tax basis. Conversely, some homeowners who are not equity-, credit-, or income-constrained choose not to exercise options that appear to be in the money. There are several possible reasons for such behavior. For instance, a homeowner who expected to move in the near future might not have enough time to recoup the transaction costs of refinancing.

In our model of refinancing, the dependent variable is a discrete binary indicator that assumes the **value** of 1 when the homeowner refinances and zero otherwise. We use logit **analysis** to estimate the effect of various explanatory variables on the probability that a loan will be refinanced. The explanatory variables may be categorized as (1) market interest rates and other factors in the lending environment affecting the cost, both **financial** and nonfinancial, of carrying out a refinancing transaction, (2) the credit history of the homeowner, and (3) an estimate of the post-origination LTV ratio. In addition, as in most prepayment models, we include the number of months since origination (or the "age" of the mortgage) to capture age-correlated effects not stemming from equity, credit, or the other explanatory variables. (See the appendix for further explanation of logit **analysis** and how it is applied in this case.) More details on the definition and specification of these variables follow; Table 1 presents summary statistics.

THE INCENTIVE TO REFINANCE

Theory suggests that homeowners will refinance if the benefits of doing so-that is, the reduction in after-tax mortgage interest payments over the expected life of the loan-exceed the transaction costs of obtaining a new loan. Accordingly, measuring the strength of the incentive to refinance involves a comparison of the contract rate on the existing mortgage with the rate that could be obtained on a new mortgage. In addition, account should be taken of transaction costs (such as discount points and assorted closing costs), the opportunity cost of the time spent shopping for and qualifying for a new loan, and interest rate volatility, which influences the **value** of the call option.5

(Table Omitted)

Captioned as: Table 1

There are many ways to measure the strength of the incentive to refinance, none of which is perfect (see, for example, Richard and Roll 1989]). In this study, we employ the simplest of them-the spread between the contract rate on the existing loan (C) and the prevailing market rate (R), that is:

SPREAD = C-,

where (t) represents the time period. For all observations in our sample, C is the Freddie Mac national average commitment (contract) rate on fixed-rate loans for the month in which the existing loan closed.6 This is the so-called A-paper rate, or the rate available to the best credit risks. Likewise, for those homeowners who did refinance, R is also the national average A-paper contract rate for the month in which the new loan closed.

While SPREAD is a simple measure and tends to represent the way homeowners think about the refinancing decision, it has some drawbacks. First, it does not explicitly account for transaction costs, which are likely to vary across borrowers and over time. However, one could imagine that transaction costs create an implicit critical threshold of SPREAD, say 100 to 150 basis points, that must be exceeded to trigger a refinancing. Another drawback of SPREAD is that it does not take into account the fact that the **financial** benefit of refinancing is a function of the expected life of the new loan. However, experimentation with alternative measures that do explicitly account for transaction costs and holding period revealed that the effects of creditworthiness and home equity on the probability that a loan will be refinanced are insensitive to the measure employed.7

An important issue that arises when using SPREAD in cross-sectional **analysis** is the assignment of the **value** of R to those individuals who did not refinance. Several possible approaches exist for assigning a **value**, and there is a certain amount of arbitrariness in selecting any particular one.8 In tackling this problem, we noted that those who did refinance rarely did so at the largest spread (the lowest **value** of R) that occurred over the period from their original purchase to the date they refinanced (Chart 3). If all the values of SPREAD observed over that period were ranked from highest to lowest, on average those who refinanced did so at about the seventy-fifth percentile. Accordingly, we assigned nonrefinancers the **value** of R associated with the seventy-fifth percentile of spreads observed over the period from the date of original purchase to the end of our sample period (December 1994).

Note that by basing C and R on the A-paper rate, we explicitly excluded from SPREAD any influences that individual borrower characteristics might have on the actual values of particular individuals. The effects of those individual characteristics are captured by the credit and equity variables, as well as by the error term. In addition, we ignored the fact that the values of C and R for any one individual are likely to deviate somewhat from the national average because of regional differences in mortgage interest rates or differences in the shopping and bargaining skills of refinancers.

(Chart Omitted)

Captioned as: Chart 3

VOLATILITY

As noted above, standard option theory suggests that there is **value** associated with not exercising the option to refinance that is increasing with the expected future volatility of interest rates. Assuming that one can correctly measure expected future volatility, theory also suggests that, when included in a model such as ours, volatility should have a negative sign. That is, higher volatility should reduce the probability that a loan will be refinanced. The expected effect of volatility has been found in some studies on this topic. For example, Giliberto and Thibodeau (1989), who measure volatility as the variance of monthly averages of mortgage interest rates over their sample period, find that greater volatility tends to increase the age of a mortgage (and

decrease prepayments). In contrast, Caplan, Freeman, and Tracy (1993) find their measure of expected future volatility to be insignificant and drop it from their **analysis**.

Although the theoretical effect of expected future volatility on the probability that a loan will be refinanced is negative, actual volatility during a given time period should correlate positively with the probability of refinancing during that period. That is, if market interest rates during the relevant **interval** are relatively volatile, a homeowner will be more likely to observe an opportunity to refinance than if rates are relatively stable.

To capture this effect, we include as an explanatory variable the **historical** standard deviation (HSD) of market rates during the time **interval** from purchase to refinancing or from purchase to the end of the sample period. HSD is measured as the standard deviation of the ten-year Treasury bond rate. We expect this variable to be directly related to the probability that a loan will be refinanced.

LENDING ENVIRONMENT

As noted by many industry experts, between the late 1980s and the early 1990s, the mortgage lending industry became more aggressive in soliciting refinancings. To encourage refinancing, mortgage servicers began contacting customers with spreads above some threshold, often as low as 50 basis points, and informing them of the opportunity and benefits of refinancing. Transaction costs declined as competing lenders reduced points and fees (Chart 4). Indeed, many lenders began offering loans with no out-of-pocket costs to borrowers. "Psychic" transaction costs were also reduced as lenders introduced mortgage programs that minimized the **financial** documentation required of borrowers ("no doc" or "low doc" programs) and drastically shortened the periods from application to approval and from approval to closing. This change in the lending environment likely increased the probability of a loan being refinanced, all else being equal.

To capture this effect, we introduce an explanatory variable termed lending environment (LE). LE is defined as the change in the average level of points and fees (expressed as a percentage of the loan amount) on conventional fixedrate loans closed between the time of the original purchase and either refinancing or the end of the sample period.

(Chart Omitted)

Captioned as: Chan 4

PERSONAL CREDITWORTHINESS

Since credit history is a key determinant of mortgage loan approval, it clearly should have some bearing on the likelihood that a loan will be refinanced. However, because of a lack of data, this effect has never before been quantified. Our study is able to overcome this obstacle. The Mortgage Research Group (MRG)the source of most of our datahas matched complete TRW credit reports to the individual property records that make up our sample of loans (see box). Using this matched data, we are able to test our hypothesis that, other things being equal, the worse an individual's credit rating, the lower the probability that he or she will refinance a mortgage, either because the homeowner cannot qualify for a new loan or because the interest rate and transaction costs at which he or she can qualify are too high to make it financially worthwhile.

The most general measure of an individual's credit history presented in the TRW reports is the total number of "derogatories."9 A derogatory results from one of four events:

* a charge off: when a lender, after making a reasonable attempt to collect a debt, has deemed it uncollectible and has elected to declare it a bad debt loss for tax purposes. There are no hard and fast rules specifying when a lender can elect to charge off a debt or what represents a reasonable effort to collect. A charge off may result from a bankruptcy, but most often it is simply the result of persistent delinquency.

* a collection: when a lender has enlisted the services of a collection agency in an effort to collect the debt.

* a lien: a claim on property securing payment of a debt. A lien (for example, a tax lien or mechanics lien) is a public derogatory because it is effected through the courts and is a matter of public record.

* a judgment: a claim on the income and assets of an individual stemming from a civil law suit. Like a lien, a judgment is a public derogatory.

Somewhat more specific indicators of an individual's credit history are the worst now (WRSTNOW) and worst ever (WRSTEVER) summary measures across all credit lines. As the names imply, these variables capture an individual's worst payment performance across all sources of credit as of some moment in time (now) and over the individual's entire credit history (ever). At the extremes, either variable can take on a **value** of 1 (all credit lines are current) or a **value** of 400 (a debt has been charged off). Intermediate values capture the number of days a scheduled payment has been late: 30 (a scheduled payment on one or more credit lines is thirty days late), 60, 90, or 120.10 Note that a 400 constitutes a derogatory, whereas some lesser indicator of credit deterioration, such as a 90 or 120, does not.

To clarify how the WRSTNOW and WRSTEVER measures are used to assess an individual's credit status, we offer the example of a homeowner who has three credit lines-a home mortgage, a credit card, and an auto loan (Table 2). At the beginning of the homeowner's credit history (t-11), all three credit lines are current, giving the homeowner WRSTNOW and WRSTEVER values of 1. For some reason-perhaps loss of employment, illness, or divorce-this individual begins to experience some difficulty meeting scheduled payments on a timely basis. The credit card payment due becomes 120 days late in period t-7, prompting the lender to charge off that debt in period t-6, at which point both WRSTNOW and WRSTEVER take on a **value** of 400. Eventually, this individual gets all credit lines current again, bringing WRSTNOW down to 1 by period t-1. However, WRSTEVER remains at 400 because of the charge off of the credit card debt in period t-6. Indeed, once someone experiences credit difficulties, his or her credit history is likely to be affected for a long time.

(Table Omitted)

Captioned as: Table 2

We now examine a cross tabulation of the WRSTNOW and WRSTEVER values for all individuals in our sample (Table 3). For WRSTNOW, 85.5 percent of the sample have a **value** of 1 while 8.0 percent have a **value** of 400. Values from 30 to 120 represent just 6.5 percent of the total. In contrast, for WRSTEVER, 18.4 percent of the sample have a **value** of 400 while just 52.9 percent have a **value** of 1. Thus, although at any point in time nearly nine of every ten individuals have a perfect credit rating (WRSTNOW=1), at some time in their credit history roughly half the population experienced something less than a perfect credit rating (WRSTEVER> 1). In fact, 8.0 percent have a WRSTNOW of 1 but a WRSTEVER of 400.11

The ideal data set for determining the effect of credit history on the probability that a loan will be refinanced would include a credit snapshot as of the date the home was originally purchased and periodic updates, perhaps once per quarter, as the loan ages. With this information, the researcher could determine whether the homeowner's credit history had deteriorated since the purchase of the home. Unfortunately, data sets that link property transaction data with credit histories are a relatively new phenomenon, so these periodic updates of the credit history are not yet available. As a second-best alternative, we use one credit snapshot-as of the second quarter of 1995-that includes both a current (WRSTNOW) and a backward-looking (WRSTEVER) credit measure. We included these measures of creditworthiness in numerous specifications of our logit model and, regardless of specification, found that they were both statistically and economically significant in determining refinancing probability. Moreover, by comparing WRSTNOW with WRSTEVER, we were able to identify cases where a mortgagor's credit history had improved over time, and found some evidence

that improvement reduced, but did not completely overcome, the negative impact of a WRSTEVER **value** of 4002

POST-ORIGINATION HOME EQUITY

In addition to a poor credit history, another factor that could prevent a homeowner from refinancing, regardless of how far interest rates have fallen, is a decline in property **value** that significantly erodes that owner's equity. For example, if a homeowner originally made a 20 percent down payment (origination LTV ratio=80 percent), a 15 percent decline in property **value** following the date of purchase would push the post-origination LTV ratio to nearly 95 percent, typically the maximum allowable with conventional **financing**. Loan underwriters would likely be concerned that the recent downward trend in property values would continue and therefore would be reluctant to approve such a loan.

In addition, an LTV ratio exceeding 80 percent would typically require some form of mortgage insurance, which would increase transaction costs and reduce the effective interest rate spread by as much as 25 to 50 basis points. If the original LTV ratio was greater than 80 percent, correspondingly smaller declines in property **value** would have similar effects. In contrast, increases in property **value** would likely raise the probability of refinancing. Greater equity simply makes it easier for homeowners to qualify for a loan since the lender is exposed to less risk. It may also increase the incentive to refinance for homeowners who wish to take equity out of their property (known as a cash-out refinancing). Furthermore, if price appreciation substantially lowers the post-origination LTV ratio, a borrower may be able to use refinancing to reduce or eliminate the cost of mortgage insurance, thereby increasing the effective interest rate spread.

(Table Omitted)

Captioned as: Table 3

To capture the effect of changes in home equity on the probability of refinancing, we enter an estimate of the post-origination LTV ratio as an explanatory variable. The LTV ratio's numerator is the amortized balance of the original first mortgage on the property, calculated by using standard amortization formulas for fixed-rate mortgages and the interest rate assigned to that loan, as discussed above.13 The denominator is the original purchase price indexed using the Case Shiller Weiss repeat sales home price index for the county in which the property is located. While repeat sales home price indexes are not completely free of bias, they are superior to other indicators in tracking the movements in home prices over time. This approach allows us to calculate a post-origination LTV ratio for each month from the date of purchase to either the date of refinance or the end of the sample period.

For loans that were refinanced, the post-origination LTV ratio used is the estimate for the month in which the refinance loan closed. However, as in the case of interest rate R, a **value** of the post-origination LTV ratio must be assigned to those observations that did not refinance. We noted that, on average, homeowners who refinanced did so at the forty-fifth percentile of values of the LTV ratio observed from the date of purchase to the date of refinance. On the basis of this observation, the LTV ratio assigned to those who did not refinance is the average over the entire period from the date of purchase to the end of the sample period.

We should note that virtually all of the movement in the LTV ratio is the result of changes in the **value** of the home. The amount of amortization of the original balance of a mortgage is relatively modest over the typical life of the mortgages in our sample. In contrast, over the time period represented by this sample, home price movements have been quite dramatic in some regions. For example, the Case Shiller Weiss repeat sales indexes suggest that home prices in the California counties included in our sample declined by roughly 30 percent from 1990 to 1995.

AGE OR BURNOUT"

The actual prepayment performance of mortgage pools typically shows an increase in the conditional prepayment rate during roughly the first fifty

to sixty months, at which point loans are described as being "seasoned." As the aging process continues, the remaining loans in a pool become quite resistant to prepayment, even with strong incentives-a phenomenon known as burnout. To capture this effect, most prepayment studies include the age of the loan or the number of months since origination as an explanatory variable.

One explanation for burnout is that homeowners prevented from refinancing by credit, equity, and/or income constraints come to dominate mortgage pools over time as homeowners who are not similarly constrained refinance or sell their homes. To the extent that our equity and credit variables capture this effect, the age of the loan per se should be less important than it would be in a model that does not include those variables. However, recognizing that credit and equity may not capture all age-correlated effects, we also include AGE as an explanatory variable. Because the effect of aging may not be a simple linear one, we also include age squared (AGESQ). In comparing the frequency distribution of AGE for homeowners who refinanced with the corresponding distribution for homeowners who did not, we see that the general shape of these distributions is similar-although, as one would expect, the proportion of higher AGE values is greater for nonrefinancers than for refinancers (Chart 5).14

EMPIRICAL FINDINGS

Logit estimations of our model for the entire sample-that is, all regions combined-appear in Table 4. We account for the effect of credit on the probability of refinancing by dividing the sample into three subsamples: individuals with values of WRSTNOW equal to 1 (good credits), individuals with WRSTNOW between 30 and 120 (marginal credits), and individuals with WRSTNOW equal to 400 (bad credits). We then estimate our model for each of the subsamples while dropping the credit history variable. We eliminate this variable because variations in market interest rates relative to the contract rate on a homeowner's existing mortgage would have a greater effect on the refinancing probability of a borrower with a perfect credit history than on one with serious credit difficulties. This variability in responsiveness suggests that there should be significant interactions between credit history and the other explanatory variables, particularly SPREAD.

In addition, it is not clear whether the credit variables WRSTNOW and WRSTEVER should be viewed as continuous, such as crude credit scores, or as categorical.15

Our results confirm that credit history has a marked effect on the probability of refinancing. The coefficient on

SPREAD for good credits is approximately twice as large as it is for bad credits, with a corresponding sizable drop in statistical significance in the latter case. Similarly, we find that the coefficients on HSD are positive and highly significant, although slightly smaller and somewhat less significant for the WRSTNOW=400 subsample. While high values of HSD indicate more opportunities for a mortgagor's option to be in the money, such values have less impact on the refinancing probability of credit-constrained borrowers. As expected, we find that the coefficients of the variable SPREAD are uniformly significant and positive across the subsamples.

(Chart Omitted)

Captioned as: Chart 5

(Table Omitted)

Captioned as: Table 4

Changes in home equity also have an important influence on the probability of refinancing, as evidenced by the negative sign and high level of significance of the LTV ratio. We demonstrate the estimated effect of changes in house price by plotting simulated values of the probability of refinancing for different levels of the post-origination house price as a percentage of the original purchase price (Chart 6). Note that in Table 4, the coefficient on the LTV ratio is somewhat larger for the bad credit

group, suggesting that to some extent there is a trade-off between equity and credit rating.

Lending environment is also significant and bears the predicted sign, suggesting that increased lender aggressiveness and consumer **financial** savvy have boosted the probability that a loan will be refinanced. Again note that the coefficient of LE is somewhat greater for bad credits than for good credits, suggesting that an important element of increased lender aggressiveness has been the increase in subprime credit quality lending, or lending to borrowers with credit histories worse than that required in the A-paper market. Finally, AGE and AGESQ are significant with negative signs, indicating that credit and equity do not explain all of the decline in probability of refinancing as a mortgage ages.

These results emphasize the dependence of estimates of interest rate sensitivities on credit factors. Pools of mortgages with relatively high proportions of borrowers with poor credit histories will experience significantly slower prepayment speeds, all else being equal. Investors in mortgage-backed securities are affected by the credit conditions of the households represented in the underlying pools of mortgages even though they may be insulated against homeowner default per se. Moreover, our results suggest that a change in the overall lending environment has occurred over the past decade, probably because lenders have become more aggressive and borrowers more sophisticated. All else being equal, this change has increased the probability that a homeowner will refinance.

EFFECTS OF AN IMPROVEMENT IN CREDIT RATING

The summary measures of credit history used in this study suggest that the credit performance of many individuals in our sample has improved: for these individuals, WRSTNOW has a lower **value** than WRSTEVER. As Table 3 shows, 8.0 percent of the sample have a WRSTEVER of 400 (the worst credit classification) and a WRSTNOW of 1 (the best credit classification).

To investigate the extent to which improvement in a homeowner's credit history affects the probability of refinancing, we first select all those cases in which WRSTEVER is 400 (18.4 percent of the total sample). We then divide that group into three subsamples based on the extent of improvement: WRSTEVER=40Q, WRSTNOW=1; WRSIEVER=400, 1<WRSTNOW<400; and WRSIEVER=400, WRSTNOW=400. Next we estimate our model, absent the credit history variable, over these three subsamples. We find that the coefficients on SPREAD and HSD are larger for the subsample with the greatest improvement than for the subsample with no improvement. These results provide some support for the hypothesis that improvement in one's credit rating increases the probability of refinancing (Table 5).

SIMULATING THE EFFECTS OF CREDIT AND EQUITY ON THE PROBABILITY OF REFINANCING

Using the separately estimated equations for the WRSTNOW= 1 and WRSTNOW= 400 subsamples, we simulate values for the probability of refinancing for hypothetical individuals with different credit histories and different values of the post-origination LTV ratio (Table 6). The four columns of this table represent alternative combinations of the variables WRSTNOW and the LTV ratio. Moving down each column, we see that the variable SPREAD rises from 0 to 300 basis points, an increase that should normally motivate refinancing. The first column, with WRSTNOW= 1 and the post-origination LTV ratio=60 percent, shows how an individual who is neither equity- nor credit-constrained would react to an increase in SPREAD. Note that with SPREAD=0, the probability of refinancing is 0.29, suggesting that refinancings motivated by the desire to extract equity from the property are fairly high among this group. As SPREAD rises to 300 basis points, the probability of refinancing essentially doubles, reaching nearly 60 percent. In the second column, where the LTV ratio=100 percent, the probabilities drop sharply; at SPREAD=0, the probability is just 0.1, while at SPREAD=300, the probability is 0.32, about half of that when the LTV ratio=60 percent.

(Chart Omitted)

Captioned as: Chart 6

In contrast, the third and fourth columns depict an individual who is severely credit-constrained (WRSTNOW=400). As suggested above, having substantial equity can overcome many of the problems associated with a poor credit history, particularly because more lenders have moved into subprime lending programs. With the LTV ratio=60 percent, probabilities of refinancing are essentially the same at SPREAD=0 and SPREAD=100 as in the WRSTNOW=1 case. However, without substantial equity (an LTV ratio=100 percent), the probability of refinancing is not only low but also unresponsive to increases in SPREAD.

(Table Omitted)

Captioned as: Table 5

(Table Omitted)

Captioned as: Table 6

Additional simulations test the marginal effect on the probability of refinancing of relevant changes in the model's other explanatory variables (Table 7). We saw in Table 1 that the mean **value** for LE for refinancers is 24 basis points. The results reported in Table 7 indicate that, all else being equal, this mean **value** of LE results in a 0.2 increase in the probability of refinancing. Comparing Table 7 with Table 6, we conclude that the change in the lending environment over the past decade has had an effect on the probability of refinancing equivalent to moving from an LTV ratio of 100 percent to an LTV ratio of 60 percent-a very powerful effect. Similarly, each year in which a loan ages reduces the probability of refinancing by 0.1, all else being equal.

CONCLUSION

Our **analysis** provides compelling evidence that a poor credit history significantly reduces the probability that a homeowner will refinance a mortgage, even when the **financial** incentive for doing so appears strong. Moreover, consistent with previous studies, we find that refinancing probabilities are quite sensitive to the amount of equity a homeowner has in his or her property. Homeowners with poor credit histories and low equity positions cannot easily meet lenders' underwriting criteria, so they are often blocked from obtaining the replacement **financing** necessary to prepay their existing mortgage.

On another level, this research contributes to the evidence that households' **financial** conditions can have significant effects on the channels through which declines in interest rates influence the overall economy. From the broadest viewpoint, mortgage refinancings can be viewed as redistributions of cash flows among households or investment intermediaries. For those households able to reduce costs by locking in a lower interest rate on their mortgage, refinancing is likely to have a wealth or permanent income effect that might boost overall consumption spending. Conversely, to the extent that households are unable to obtain replacement **financing** at lower interest rates because of deteriorated credit histories or erosion of equity, the stimulative effect on consumption would likely be less.

Of course, refinancing decisions also affect the investors in the various cash flows generated by pools of mortgages. When homeowners refinance, those investors lose above-market-rate income streams and so are keenly interested in any factors that may have a significant bearing on the probability of refinancing. This **analysis** demonstrates that, in addition to monitoring changes in interest rates and home prices, those investors should be concerned with the credit histories of the homeowners represented in a particular pool of mortgages as well as trends in those credit histories over time. Despite guarantees against credit risk, the relative proportions of credit-constrained households represented in pools of mortgages will have a significant impact on the prepayment behavior of those pools under various interest rate and home price scenarios.

(Table Omitted)

Captioned as: Table 7

(Table Omitted)

Captioned as: APPENDIX: MODELING THE DECISION TO REFINANCE

Footnote:

ENDNOTES

Footnote:

Stavros Peristiani, Paul Bennett, and Richard Peach are economists at the Federal Reserve Bank of New York. Gordon Monsen is a managing director in Asset Trading and **Finance** and Jonathan Raiff is a first vice president in Mortgage Strategy at PaineWebber Incorporated. The authors wish to thank Elizabeth Reynolds for outstanding technical support on this paper.

Footnote:

1. Another factor that may have impeded a borrower's ability to refinance is a decline in household income. Unfortunately, the data set used in this study does not include information on an individual borrower's income at the time of the initial purchase of the home or afterward.

2. In the literature on this topic, a distinction is made between the values of LTV ratios, income, and credit history at the time the mortgage loan is originated (the origination values) and the values of those variables at some point in time after the origination (the post-origination values). The post-origination values are the most relevant for the decision to prepay a mortgage, but they also tend to be the most difficult on which to obtain data.

3. Homeowners' assessments of the current market values of their properties may be biased, particularly during periods when there are significant changes in those values. See, for example, DiPasquale and Sommerville (1995) and Goodman and Ittner (1992).

Footnote:

4. Case Shiller Weiss, Inc., of Cambridge, Massachusetts, provided these home price indexes.

5. See Follain, Scott, and Yang (1992) and Follain and Tzang (1988).

6. The interest rate on existing loans C is not directly observed in the data base. An estimate of that interest rate can be derived from information on the original loan balance, original maturity, and periodic readings of the amortized balance, which is reported in the TRW credit reports discussed below.

Strictly speaking, an **interval** of thirty to sixty days usually separates the date of application for a mortgage from the date of closing, although borrowers typically have the option of locking in the interest rate at the time of application or letting the rate float, in some cases up to the date of closing. We experimented with lagging the national average mortgage interest rate by one and then two months and found that in neither case were the results significantly different from those we obtained using the average rate for the month in which the loan closed.

7. In a more technical version of this study, we tested four alternative, increasingly complex measures of the incentive to refinance. Details on

Footnote:

the definitions and specifications of these measures, as well as the estimation results, are presented in Peristiani et al. (1996).

8. For example, Archer, Ling, and McGill (1995) assign to those observations that did not refinance the lowest monthly average Freddie Mac commitment rate on thirty-year fixed-rate mortgages over the two-year time **interval** of their study.

9. In the technical version of this study (Peristiani et al. 1996), we use total derogatories as an explanatory variabLe in determining the probability of refinancing and find it to be highly significant with the predicted sign, although somewhat less significant than WRSTNOW or WRSTEVER.

10. In fact, each variable can take on more values than those listed. For example, a **value** of 34 indicates that an individual is persistently thirty days late. For the purposes of this study, we have constrained WRSTNOW and WRSTEVER to take on only those values cited in the text.

11. To an increasing extent, mortgage lenders are relying on a single credit score summarizing the vast amount of information on an individual's credit report. For an overview of this issue, see Avery, Bostic, Calem, and Canner (1996). As an extension of the research on the effect of credit histories on mortgage refinancings, credit scores could also be tested as an alternative measure of creditworthiness.

12. For additional information on these alternative specifications, see Peristiani et al. (1996).

Footnote:

13. The presence of second mortgages and home equity loans introduces additional considerations into the issue of refinancing. On the one hand, second mortgages and home equity loans would tend to reduce a homeowner's equity. On the other hand, since second mortgages and home equity loans typically have interest rates well above the rates on first mortgage loans, the spread based on the homeowner's weighted-average cost of credit would likely be higher. Although the MRG data base indicates the presence and amount of second mortgages and home equity loans taken out since the original purchase, we do not investigate their effect on refinancing probabilities. This is an area for future research.

14. As noted earlier, the sample excludes observations with AGE of less than twelve months.

15. Dividing the sample into three subsamples based on credit rating is equivalent to estimating the model over the entire sample with dummy variables for the three credit classifications and fully interacting those dummy variables with the other explanatory variables of the model.

Reference:

REFERENCES

Reference:

Archer, Wayne, David Ling, and Gary McGill. 1995. "The Effect of Income and Collateral Constraints on Residential Mortgage Terminations." National Bureau of Economic Research Working Paper no. 5180, July.

Avery, Robert V., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. 1996. "Credit Risk, Credit Scoring, and the Performance of Home Mortgages." FEDERAL RESERVE BULLETIN, July: 621-48.

Bernanke, Ben. 1993. "Credit in the Macroeconomy." Federal Reserve Bank of New York QUARTERLY REVIEW 18, no. 1: 50-70.

Caplin, Andrew, Charles Freeman, and Joseph Tracy. 1993. "Collateral Damage: How Refinancing Constraints Exacerbate Regional Recessions." National Bureau of Economic Research Working Paper no. 4531, November.

Cunningham, Donald F., and Charles A. Capone, Jr. 1990. "The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages." JOURNAL OF **FINANcE** 45, no. 5: 1687-703.

Reference:

DiPasquale, Denise, and C. Tsuriel Soierville. 1995. "Do House Price Indexes Based on Transacting Units Represent the Entire Stock? Evidence from the American Housing Survey." JOURNAL OF HOUSING ECONOMICS 4, no. 3: 195-229.

Estrella, Arturo. 1997. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." Federal Reserve Bank of New York Research Paper no. 9716. Forthcoming in JoURNAL OF BUSINESS AND ECONOMIC STATISTICS.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Peterson. 1988. "**Financing** Constraints and Corporate Investment." BROOKINGS PAPERS ON ECONOMIC ACTIVITY, no. 1: 141-95.

Follain, James R., James 0. Scott, and TL Tyler Yang. 1992. "Microfoundations of a Mortgage Prepayment Function." JOURNAL OF REAL ESTATE AND ECONoMICS 5, no. 2: 197-217.

Reference:

Follain, James R, and Dah-Nein Tzang. 1988. "Interest Rate Differential and Refinancing a Home Mortgage." APPRAISAL JOURNAL 56, no. 2: 243-51.

Giliberto, S. Michael, and Thomas G. Thibodeau. 1989. "Modeling Conventional Residential Mortgage Refinancings." JouRNAL OF REAL ESTATE

FINANCE AND ECONOMICS 2, no. 4: 285-99.
Goodman, John L., and John B. Ittner. 1992. "The Accuracy of Home Owners Estimates of House **Value**." JOURNAL OF HOUSING ECONOMICS 2, no. 4: 339-57.
Green, William H. 1993. ECONOMETRIC **ANALYSIS**. 2d ed. New York: MacMillan Publishing Company.
Reference:
Maddala, G.S 1983. LIMITED-DEPENDENT AND QUALITATIVE VARIABLES IN ECONOMETRICS. Cambridge: Cambridge University Press.
Peristiani, Stavros, Paul Bennett, Gordon Monsen, Richard Peach, and Jonathan Raiff 1996. "Effects of Household Creditworthiness on Mortgage Refinancings." Federal Reserve Bank of New York Research Paper no. 9622, August.
Richard, Scott F., and Richard Roll. 1989. "Prepayments on Fixed-Rate Mortgage-backed Securities." JOURNAL OF **PORTFOLIO** MANAGEMENT 15, no. 3: 73-82.
Schorin, Charles N. 1992. "Modeling and Projecting MBS Prepayments." In Frank J. Fabbozi, ed., HANDBOOK OF MORTGAGE-BACKED SECURITIES. Chicago: Probus Publishing Company.

Special Features: Charts; Graphs; Appendix; Equations; References
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 9140 (CN=Statistical data); 8120 (CN=Retail banking); 3400 (CN=Investment analysis)
Descriptors: Studies; Mortgages; Consumer credit; Effects; Refinancing; Statistical data
Geographic Names: US

36/9/10 (Item 10 from file: 268)
00301644
Report card on **value** at risk: High potential but slow starter
Beder, Tanya Styblo
Bank Accounting & Finance , v 10 , n 1 , p 14-25 , Fall 1996 **Document Type:** Journal Article **ISSN:** 0894-3958 **Journal Code:** ABAF **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER: K
Word Count: 06177

Abstract:

There are more questions than answers after 3 years of use of **value** at risk (VAR), although it shows increasing promise as a risk-management tool. There are 3 main types of VAR: 1. variance/covariance VAR, 2. **historical** VAR, and 3. simulation VAR. Seven lessons about VAR are: 1. For instruments with nonlinear price functions, variable/covariance VAR understates risk. 2. **Historical** VAR and simulation VAR can differ drastically. 3. Mapping can impair VAR calculations. 4. Poor assumptions about diversification can lead to flawed results. 5. Combining adjusted VARs from different time periods can be misleading. 6. VARs may be less comparable than they appear. 7. Accounting and economic measures may not mix. Once the outlook preferences and the complexity of the **portfolio** are analyzed and one or more VARs are selected, users must make decisions about several important dimensions of the calculation:

1. the length of the VAR horizon, 2. database, 3. correlation assumptions, and 4. mathematical engine and quantitative approach.

Text:

Headnote:

For all its cachet, **value** at risk is no cookie-cutter solution to the risk-management problems facing **financial** institutions. Implementation is hampered by the need to make simplifying assumptions. Differences in VAR methodology and the many assumptions required mean that risk managers should have a clear understanding of the components of their VAR measures and must combine VAR with other risk tools.

Vile **value** at risk (VAR) shows increasing promise as a risk-measurement tool, there are more questions than answers after three years of use. Despite this, the concept is widely endorsed by regulators such as the Bank for International Settlements, the Federal Reserve, the Office of the Comptroller of the Currency, and the Securities and Exchange Commission. It is mandated for many under generally accepted accounting principles, is part of the rating process by agencies, and is encouraged by key industry groups such as the Group of Thirty, the Derivatives Policy Group, and the International Swaps and Derivatives Association. But implementation of VAR is harder than grasping the simplicity of its concept. First, not all VARs are equal. Second, vast quantities of data and significant modeling or systems efforts may be required. Third, firms must design and implement risk-management add-ons to address VAR's limitations and weaknesses. While dealers typically are further along with VAR implementation than end users, few if any are finished with the process. This article surveys the current realities of VAR and what we have learned to date.

Three Main Types of VAR

VAR is the great equalizer. It translates the risk of any **financial** instrument into its potential loss under specific assumptions.1 There are three main types of VAR: variance/covariance VAR, **historical** VAR, and simulation VAR.

Variance/covariance VAR

Under this method, sometimes called "analytic VAR," **financial** instruments are decomposed (or "mapped") into delta equivalents2 consisting of basic **financial** building blocks, or market factors. Once **historical** or other distributions for these market factors are specified, VAR and other measures are computed using standard statistical techniques. In most cases, **historical** data is used to build the variance/covariance matrix for the market factors, making this aspect of the calculation dependent upon the time period selected. Over the past two years, data sets that provide distributions for many common market factors (for example RiskMetrics3) have become available, as have commercial software packages that perform VAR computations.

Historical VAR

Under this method, **financial** instruments are analyzed over the number of days in the **historical** observation period (for example, 100 days), and the actual change that was experienced in the **value** of each **financial** instrument is calculated using the desired time horizon (for example, overnight). Note that while most users analyze **financial** instruments specifically, some translate their **financial** instruments into "equivalent" building blocks or market factors and calculate the changes on these. Once the changes in **value** are calculated, each change is added to today's **value** for the **financial** instrument or its

"equivalent" to produce an array of observations. As this replicates **historical** behavior, the risk view depends upon the time period selected. To complete the calculation, the array is analyzed statistically. For example, if there are 100 observations, the 5th lowest observation **value** would be the one-day 95% confidence **interval** VAR.

Simulation VAR

Under this methodology, the theoretical probability distribution of changes in **value** for each **financial** instrument or its "equivalent" is calculated for the desired time horizon (for example, over two weeks) as per the distribution parameters specified in the simulation. Typically, correlations and lognormal or other distributions are incorporated. The theoretical changes in values are then added to today's **value** for the **financial** instrument or its "equivalent" and arrayed as in the case of **historical** VAR to produce the desired confidence **interval** VAR. The process is often completed under varying sets of parameters.

Each type of VAR has its strengths and weaknesses. Variance/covariance VAR is the least computationally intensive and free data is available. However, it is based on normal or lognormal distributions so it misses fat-tailed behavior4 and does not properly incorporate options or other nonlinear instruments. **Historical** VAR is the easiest to implement from a systems perspective and may be the easiest to explain to the nonmathematically inclined. However; its output depends heavily on the time period selected (simply stated, history must repeat itself). Simulation VAR can incorporate any joint distribution for the market factors, so offers the greatest flexibility for sensitivity analyses regarding market plus model issues, and fully captures nonlinear instruments. However, it has the greatest systems, programming, and data needs.

Seven Lessons about VAR

Beginning in 1994, dealers focused on implementing at least some VAR measure and devoted their resources to data, systems, and programming challenges. Risk-management software vendors took a similar approach, focusing primarily on the need to expand their systems to include at least one VAR alternative. At first, most implemented variance/covariance or **historical** VAR calculations. Larger corporations implemented VAR as well, with the goal of comparing the treasury area's performance versus an established internal benchmark. Some institutional investors and investment managers (particularly insurance companies, mutual funds, and "manager of managers") began to implement VAR over the past 6 to 12 months, with the goal of calculating riskadjusted **portfolio** performance. Many smaller corporations, as well as pension funds, public funds, foundations, and endowments, have started to address VAR more recently.

To date and in general, the theoretical discussions of VAR far exceeded firms' actual practices.5 This is due to the many practical issues that complicate and surround its implementation. However, valuable lessons have been learned, and these are being addressed as VAR approaches its third year of use in risk management. Seven lessons follow.

Lesson One: For instruments with nonlinear price functions, variance/covariance VAR understates risk

The variance/covariance approach significantly understates risk for portfolios with options or **financial** instruments with nonlinear price functions,6 particularly during periods of large volatility or with large changes in the price of the underlying. Most dealers with significant nonlinear exposures have implemented or are switching over to simulation-based VAR calculations for at least the nonlinear books within their businesses. This presents aggregation issues regarding VARs calculated with different methods over different time horizons. Research is under way regarding risk-management addons to a variance/covariance approach that better reflect nonlinear risks.

Lesson Two: **Historical** VAR and simulation VAR can differ

drastically

The **historical** VAR and simulation VAR approaches may produce vastly different results, especially when the **historical** period comprises a heavily trending market. This is due to the fact that the key variables in simulation VAR are computed according to the user's expectations or may be computed **randomly** and often differ substantially from those for the recent **historical** period. There are many types of simulation, each determined by the user's preferences and parameters. Monte Carlo simulations are the most common type of **random** simulations. To the degree **random** or user-specified expectations vary from trending market expectations, differences between the two approaches will be magnified. Note that the choice of simulation parameters is itself an important determinant of the VAR result, so some dealers and end users are beginning to stress-test the sensitivity of the VAR result to alternate sets of parameters. Appropriate stress tests vary and depend upon factors such as **portfolio** composition, holding period, risk appetite, systems capabilities, etc.

Lesson Three: Mapping can impair VAR calculations

For large dealers and end users, **historical** VAR and simulation VAR7 require vast quantities of data plus numerous pricing models. To enable calculation of VAR as models and databases are built or to reduce the total amount required, most VAR users have resorted to some degree of mapping **financial** instruments into equivalents and/or matrix pricing. This often results in significant differences between the risk/reward profile of the actual **financial** instrument and its mapped equivalent. Research is under way to learn the degree to which this impacts the VAR result, particularly in the case of nondiversified portfolios, heavily engineered instruments, exotic instruments, etc. I have reviewed several cases in which the VAR calculation was performed correctly, but the accuracy lost through mapping or matrix pricing produced misleading results for the actual **portfolio**.

Lesson Four: Poor assumptions about diversification can lead to flawed results

The variance/covariance approach requires mapping **financial** instruments into market factors that are contained in the matrix. To facilitate this process, entire instrument classes are often mapped into market indices. For example, all domestic stocks may be mapped into the S&P 500 or all corporate bonds into a swap index. For several portfolios we have reviewed, mapping an undiversified **portfolio** into an assumed diversified **portfolio** produced misleading results. Research is under way to analyze the relationship between the quality of the VAR result after such mapping and varying degrees of diversification.

Lesson Five: Combining adjusted VARs from different time periods can be misleading

Many VAR users employ different time horizons for different trading areas or asset classes. For example, an overnight horizon is used for the forward foreign exchange positions, while a longer time horizon is used for real estate or illiquid/exotic **financial** instruments. To obtain a firmwide VAR statistic for a comparable time period, adjustments are made using statistical approximations such as the square root of time. To the degree that markets do not follow linear price behavior and normal distributions (most markets do not) and to the degree that drift should be considered, misleading results will be produced by such approximations.

Lesson Six VARs may be less comparable than they appear

Performance measurement and capital allocation are common goals of VAR users. The desire is to allocate capital to areas that have the greatest performance with the least amount of risk. However, many **financial** instruments and markets are inefficient and have risk profiles that change over time. Thus, the VAR for highly liquid, diversified portfolios may be compared to the VAR for highly illiquid, undiversified portfolios, and results are often not comparable.

Furthermore, two portfolios or business areas with equivalent VAR and return may have different risk tails, thus producing different expectations of loss outside of the confidence bands. Research is under way to see what can be learned from analyzing the changes in VAR over time (that is, the first derivative with respect to time). Other research is studying the relationship between downside risk and the degree of diversification to determine how these risk dimensions should be incorporated into performance measurement and the capital allocation decision.

Lesson Seven: Accounting and economic measures may not mix Many corporations use VAR in conjunction with a benchmark in the treasury area. For many, the goal is to manage the volatility of earnings. Two common problems arise with this approach. First, accounting realities may differ significantly from economic realities. To the degree that the benchmark is accounting based and the VAR calculation is economic based, this problem will be exacerbated. Second, earnings occur continuously and involve all business activities of the company, while VAR typically is based on a snapshot of selected activities of the corporation at a point in time. Both require adjustments in how VAR is employed.

Which VAR Should You Use?

VAR research to date primarily has involved portfolios of simple, highly liquid **financial** instruments such as Treasury strips, equity index options, and forward foreign exchange contracts. Our review of dozens of dealers' and end users' risk-management techniques revealed vast differences not only in the type of VAR calculation but also in the VAR statistics produced. Variances in the VAR statistic ranged by as much as 14 times for the same **portfolio**, depending on the type of VAR calculation and the time horizon.8 Large variances in VAR have been corroborated by others' research, particularly for portfolios that contain options.9 Yet other research suggests that variances in VAR may be less significant for portfolios that do not contain options or other instruments with nonlinear price behavior, especially over one-day holding periods,lo and that the length of sampling periods plays an important role.ll

Which VAR methodology to select depends on several factors. Typically, dealers and end users with complex portfolios set a goal of implementing a consistent, firmwide VAR that reflects their outlook preferences and the complexity of the **portfolio**. For portfolios with options or significant nonlinear price behavior, the **historical** VAR and simulation VAR produce superior results to the variance/covariance VAR. However, the systems, model, data, personnel, educational, and time requirements of the **historical** and simulation VARs often result in the use of variance/covariance VARs or multiple VAR methodologies on an interim basis. The choice between **historical** and simulation VAR resides largely with the user's outlook preferences and the desire to perform sensitivity analyses. **Historical** VAR is based on actual, past market experience whereas simulation VAR is based on the user's outlook and expectations. Full sensitivity analyses can be performed only on the latter.

Once the outlook preferences and the complexity of the **portfolio** are analyzed and one or more VARs are selected, users must make decisions about several important dimensions of the calculation:

the length of the VAR horizon (overnight, two weeks, longer),
database,
correlation assumptions,
mathematical engine and quantitative approach,
percentage of outcomes to be considered,
other risk-management and risk-measurement tools combined with VAR.

The length of the VAR horizon (overnight, two weeks, longer) VAR requires the firm to select a time horizon for analyzing risk in the context of expected losses. For example, dealers often select overnight time horizons, while pension funds and corporations often select longer horizons. One challenge in the selection of the time horizon is that while a model may produce adequate views of capital at risk on an overnight or

weekly basis, it may produce inadequate risk views over time horizons of several months, a year, or longer. For example, the calculation of one-day or overnight VAR may be misleading for customized or exotic products that cannot be analyzed, action decided upon, and liquidated in such a time frame. The 1995 Basle Amendment suggests that firms employ a single time horizon of two weeks (10 business days) for VAR calculations. This may be short relative to the life of many asset classes and other exposures and potentially too long for highly liquid instruments.

A second challenge is that while longer time horizons may be preferred for instruments such as illiquid, path-dependent options, some mathematical functions are inaccurate beyond small market moves. For example, many mathematical models are incapable of handling discontinuities such as market gapping or require linearity to produce accurate information, yet these are used in pricing models that are part of the VAR calculation. Over the past two years, dozens of dealers and end users announced losses due to differences between estimated short-term profits and actual experience over longer time horizons. This suggests that firms should test the sensitivity of the VAR calculation to alternate assumptions regarding pricing models (see "Mathematical engine and quantitative approach," below) and time horizon.

For some firms, a third challenge is to compare and combine VARs calculated over alternate time frames and under different methods. As discussed above, the translation of longhorizon VARs into short-horizon VARs (and vice versa) typically assumes linearity, joint normal relationships (that is, that the square root of time is sufficient), or static relationships (that is, no drift), which may produce misleading results.

Database

VAR requires data covering all relevant market factors and variables on which to perform the calculations. Vastly different risk views may be produced by alternate data sets. For example, during a recent 24-hour period, the 10-year U.S. Treasury traded at as high a price as 103 for three hours but only at par for one hour. Thus, time of day (or intraday data versus end-of-day data) can produce contrary risk views via VAR. Different risk views can also be created by the use of **historical** versus market-implied data. Note that **historical** end-of-day data is most often employed to calculate VAR, but the **historical** period selected varies significantly from firm to firm. Some firms employ the most recent 90day time horizon while others use the past year at a minimum. Other firms expand the time horizon to capture periods of stressful market moves such as market crashes or dislocations. The proposed Basle Amendment suggests that firms employ a one-year minimum data set for VAR calculations.

Length of time is not the sole criterion to establish and test regarding the data set. As discussed, mapping procedures are a critical part of most VAR processes. Furthermore, sampling frequency and independence of data also can affect VAR significantly. For example, a one-year database comprised of 12 end-of-month data points may be no more relevant than a data set of 12 points selected through **random** chance. Alternately, theoretical mark-to-model prices for customized or illiquid instruments may be far from market prices at the time of transactions. Such data issues can cause unpleasant surprises, as experienced in 1994 by many mutual funds, pension funds, and municipalities that monitored engineered mortgage securities and/or inverse floaters at month-end based on theoretical values.12

Another decision regarding the data is whether to exclude certain data points. For example, should the data set include outlier events caused by onetime events, market gapping, or other dislocations? Such occurrences are often characterized as extreme but lowprobability events. Recent examples are the devaluation of the Mexican peso, the 1987 stock market crashes, and commodity volatility during the Gulf War. Note that two databases, distinguished by inclusion of outlier events, are likely to produce

different VAR calculations.

Yet another challenge is to determine whether an outlier event is an indication of structural change in the market. For example, fundamental change in the prepayment patterns for mortgage-based securities in the United States occurred over the past few years, driven by mortgage broker activity and education of the home owner. Before the change, conventional wisdom dictated that a drop in interest rates had to prevail for two to three months before refinancing occurred. Subsequently, this refinancing lag shortened from months to weeks, and the mortgage market demonstrated new prepayment patterns during the rally that ended with the Federal Reserve's interest-rate hike in February 1994. Thus, use of **historical** prepayment data was misleading in predicting the expected life (and therefore return) of many mortgage securities.

Some firms employ data sets based on implied market information to reduce dependence on **historical** data. Whatever the data set, firms should stress-test the sensitivity of the VAR calculation not only to exclude any data points but also for sampling error and the use of specific **historical** periods and/or mark-to-model dependence. The goal is to determine whether alternate data sets drive large differences in the **value** of VAR for the same **portfolio** or exposures.

Correlation assumptions

VAR requires that the user decide which exposures are allowed to offset each other and by how much. For example, is the Japanese yen correlated to movements in the Italian lira or the Mexican peso? Is the price of Saudi Light correlated to movements in the price of natural gas? If so, by how much? VAR requires that the user determine correlations not only within markets (for example, U.S. dollar [USD] currency underlyings vs. USD commodity underlyings) but also across markets (for example, how do changes in the bond market in the United States relate to changes in the equity market in Australia?). Note that mapping procedures have additional embedded correlation assumptions. For example, mapping individual stocks into the S&P 500 or fixed-income securities into the swap curve translate into the assumption that individual **financial** instruments move as the market overall. While this may be a reasonable assumption for well-diversified portfolios, it may not be reasonable for undiversified or illiquid portfolios.

Dealers, end users, regulators, and **financial** theorists espouse wildly different views on the topic of correlation relationships both within and across markets. For instance, pension funds have tackled correlation issues for decades in analyzing strategic versus tactical allocation of assets. Pension funds with a lack of diversification across asset classes (for example, stocks versus bonds) or capital markets (for example, domestic versus foreign) may well be considered to be in violation of the prudent man standard of the Employee Retirement Income Security Act of 1974 (ERISA). **Financial** theory133 demonstrated the **value** of diversification, both within and across markets, decades ago. While cross-border legal and netting risks may exist, these risks typically are managed and reserves are taken separately from market risks. Despite the use of separate reserves and risk calculations, the 1995 Basle Amendment allows only the extreme position of correlation within asset classes. For calculating VAR, the amendment assumes a correlation of 1 between long positions and a correlation of -1 between long and short positions. While this may be of little consequence for some relationships (for example, the correlation between strong currencies and interest rates in European Community countries), it is of huge consequence for others (for example, the correlation between the price of a restaurant stock in Sri Lanka and a Yankee bond issued by the Canadian telephone company). Not surprisingly, the rigid correlation methodology in the 1995 Basle Amendment raises VAR significantly relative to more common correlation assumptions.14

Additional challenges exist. What happens if a market breaks through its **historical** or implied trading pattern and violates the correlation assumption in place? A recent example is provided by the many

currencies that previously displayed little or no **historical** correlation to the Mexican peso but made sympathy moves during the peso's devaluation. What happens if some temporary phenomenon alters correlations significantly? For example, barrier options on spreads (also known as knock-out or knockin options) have been blamed for unexpected, high correlations during periods that market levels approach strike levels, with both the writers and the buyers of the barriers suspected of large trading volume to influence the outcome in their favor.

In CMRA's review of different approaches to VAR, some firms assumed that all cash flows were correlated across all markets, while others assumed a lower degree of correlation. Sophisticated mean-variance models, for example the one used to compute the RiskMetrics data set, allow correlation for all instruments across all markets that are covered. At the other extreme are models such as the 1995 Basle Amendment, which require correlation of 1 or -1, depending on what is least favorable to the VAR calculation.

Mathematical engine and quantitative approach All VAR calculations require the use of mathematical models to **value** individual instruments (or their components or assumed equivalents) as well as to **value** the aggregate **portfolio**. **Valuation** variances produced by widely accepted models (termed "mark-to-model" risk) are well documented and the subject of numerous articles.15 For example, the Black-Scholes versus Hull and White options models can produce differences of 5% or more in pricing, even when all input data and curve construction (that is, crossover from futures to cash, interpolation, extrapolation, etc.) are identical. In addition, the selection of probability distribution(s)16 in one model versus another varies from firm to firm and is a topic of great debate among theoreticians and practitioners alike.

While many dealers and end users are well versed in testing the behavior of an individual position or **portfolio** given market moves (for example, what happens if interest rates rise or fall by 1 basis point or by 200 basis points?), they have only recently commenced testing the behavior of individual positions or portfolios for changes in model assumptions. Given the increased pace of losses due to model risk (the risk that the market price will be different than that calculated theoretically by a model), firms should test the sensitivity of the VAR calculation to alternate mapping and model assumptions. The goal is to determine how much the risk picture changes if one changes either the underlying mathematical model or one or more assumptions regarding the data source, time of collection, curve creation, probability distribution, mathematical process, or other factors to reflect the VAR approaches described in the 1995 Basle Amendment, the RiskMetrics Technical Document, or other common VAR models. To the degree that other common models indicate an aggressive stance by the firm, an adjustment to the VAR calculation may be appropriate or a higher VAR factor may be appropriate to protect the firm's capital from a market-risk perspective. Such model-risk adjustments should be taken in addition to those for credit risk, market risk, liquidity risk, operations risk, or other standard risk reserves.

Percentage of outcomes to be considered

The VAR methodology requires the firm to select the percentage of outcomes that will be used to determine the expectation of loss. For example, some firms calculate VAR under the requirement that the outcome or a worse outcome is expected approximately 1% of the time (often called a "99% confidence **interval**"). Others pose a lower requirement of expecting the outcome approximately 10% or 5% of the time. Perhaps due to the "confidence **interval**" terminology, some firms make the mistake of equating their VAR expectation to a certainty that the firm will not lose more than the stated amount. This is incorrect.

An important challenge in selecting the percentage of outcomes is to address the firm's need for an absolute loss limit. For example, a 95% "confidence" **interval** dictates that losses are expected to

exceed the VAR limit at least once every three weeks. Users should address how large these losses may become through stress-testing and establish limits accordingly. Furthermore, users may wish to address the potential for cumulative losses, none of which exceed the VAR limit individually, to be greater than the risk appetite of the firm. What if the amount of the VAR limit is lost continuously over contiguous time horizons (for example, daily for an entire month)?

Other risk-management and risk-measurement tools combined with VAR

Most users combine VAR with stress-testing to address questions such as "How much do I expect to lose the other 1% of the time?" As with VAR, the quality of the answer depends on the inputs, including the **financial** engineer's ability to select appropriate scenarios. Both the European currency crisis and the Gulf War demonstrated that predicting factors such as "maximum" volatility is difficult and that correlation relationships can change substantially during extreme market moves. The increasing complexity and optionality of many derivatives and engineered securities make relevant scenario selection even harder. Given such challenges, firms often resort to designing stress tests that analyze large **historical** market moves.

In CMRA's experience, portfolios do not necessarily produce their greatest losses during extreme market moves. Whether asset based or asset-plus-liability based, portfolios often possess Achilles' heels that require only small moves or changes between instruments or markets to produce significant losses. Stress-testing extreme market moves does little to reveal the greatest risk of loss for such portfolios. Furthermore, a review of a **portfolio**'s expected behavior over time may reveal that the same stress test that indicates a small impact today indicates embedded land mines with a large impact during future periods. This is particularly true of options-based portfolios that change characteristics due to time, rather than due to changes in the components of the **portfolio**. For this reason, it is paramount to employ stress-testing to reveal the following:

For market variables or model assumptions that have a high likelihood of change, what is the impact of small and large changes on VAR?

For variables or exposures considered to offset each other, how do alternate correlation assumptions affect VAR?

How wide is the variance of results produced by other common VAR approaches compared to yours?

The Mexican peso devaluation in December 1994 illustrates the difficulty in using stress-testing to analyze crises. The devaluation and subsequent market dislocation caused a 30% drop in the **value** of holdings in five days, with average losses ranging between 15% and 50%. More than 400 funds and most emerging market derivatives portfolios held TELMEX stock, so they experienced significant, unexpected losses. How should such dramatic market moves be captured by the VAR calculation or other tests? In virtually all cases, the VAR calculation considered the likelihood of occurrence minuscule (far less than a 1% expectation) when analyzing either **historical** or expected movements of the peso. In virtually all cases, firms' stress tests considered far less dramatic market moves. Today, firms remain divided about including such a low-probability event in future calculations. Firms are divided as well on the inclusion of the December 1994 peso move in **historical** data sets. In other words, the peso move is considered to be an outlier, so some firms remove it from their **historical** data sets when calculating VAR. Regardless of whether such moves are included, a valuable post mortem is to assume such an event occurred and to determine whether losses expected under VAR equal those incurred. Back-testing a firm's qualitative and quantitative risk-management approach for actual, extreme events (whether market dislocations or the actions of a rogue trader) often reveals the need to adjust reserves, increase the VAR factor, adopt additional policies/limits/controls/procedures, or expand risk calculations plus reporting.

The Mexican peso crisis was not a standalone event in terms of magnitude, suggesting the importance of such back-testing. At least one major market (not an emerging market) makes a 10 or more standard deviation move every year. For example, there have been nine Hong Kong market declines greater than 20% and two more than 50% in the past 15 years. The devaluation of the Italian lira, the stock market crashes of 1987, and the oil shocks in the 1970s are further examples of market moves far beyond the 2- to 3-standard-deviation assumption used in most VAR calculations. In the case of the 1982-87 U.S. bull market followed by stock market crash (508point plunge on October 19, 1987), within six months the markets stabilized and in less than 2 years the markets returned to precrash levels.

Many risk variables such as political risk, personnel risk, regulatory risk, phantom liquidity risk, and others are difficult or impossible to capture through quantitative techniques. Yet as demonstrated by recent, well-publicized losses, such variables can cause significant risk. For this reason, VAR must be supplemented not only with stress-testing but also with prudent checks and balances, procedures, policies, controls, limits, **random** audits, appropriate reserves, and other risk measures (Exhibit 1).

VAR in Practice

Comparing five different variance/covariance, **historical**, and simulation VARs for a hypothetical **portfolio** consisting of Treasury strips plus S&P 500 equity index contracts and options shows some of the vagaries of VAR. Exhibit 2 sets forth the **portfolio** as of May 25, 1995, comprised of long positions in 2-year and 30-year U.S. Treasury strips17 and a long position in the S&P 500 equity index contract plus long and short options on the same index. The net investment in the **portfolio** is $2 million.

Variance/covariance VAR is calculated once, using the JP Morgan RiskMetrics data set. **Historical** VAR is calculated twice, using 250-day and 100-day prior **historical** periods. Simulation VAR is calculated twice, using correlations and volatilities from the RiskMetrics data set (Simulation A) and from the 10 years prior (Simulation B). The results of the calculations appear in Exhibits 3, 4, and 5.

(Chart Omitted)

Captioned as: Exhibit 1

The actual VAR statistics are set forth in Exhibit 3 and may be interpreted as follows: under the assumptions specific to the particular VAR calculation, there is a 1% (or 5%) expectation that the **portfolio** will suffer a loss greater than or equal to the statistic shown. Thus, under the assumptions made to perform **historical** VAR over a 250-day period and assuming a two-week holding period, there is a 1% expectation of loss equal to or exceeding 1.08% of the $2 million investment in the **portfolio** (that is, a loss greater than or equal to $21,600).

The distributions for the VAR calculations are set forth in Exhibits 4 and 5. For both the 1% and 5% expectation of loss results, the alternate methods produce quite different results. Several observations may be made:

In all cases, Simulation B produces much higher expected loss levels. This is due to the fact that all four other VAR calculations depend significantly upon a more recent **historical** period, whereas Simulation B is based upon correlations and volatilities drawn from a 10-year prior period.

The 100-day and 250-day **historical** VAR calculations produce quite different downside and upside risk expectations. For example, the 1% expectation of loss for VAR in the case of the 100-day **historical** simulations is a single data point, consisting of the largest loss over a single overnight and over a single 10-day trading period. Furthermore, there is high autocorrelation in the data set. In other words, not only does a 1% probability consist of only 1 of the 100 observations, but there are only 10 distinct 10-day periods. During the

100-day and 250-day periods included in the **historical** VAR calculations, the **value** of Treasury strips largely appreciated. Had a period of rising interest rates been selected, the opposite result would have been produced. The danger in basing VAR estimates on direct **historical** observations, and over short data periods, is apparent-history must repeat itself for this method to provide an accurate expectation of future loss. The loss of the fat tails due to the nonlinearity of both the options and the Treasury positions is clear when the variance/covariance VAR distribution is compared to all other results.

VAR: Only One Aspect of Risk Management

While firms typically select a single VAR measure, it is important to determine the degree to which the answer changes under different methods. Several important dimensions of VAR are now being researched and may provide insights into adjustments that may be practicable for various methods:

the impact of time horizon;

the impact of nonlinearity;

the degree of price opacity (reverse engineering complexity, illiquid underlyings, illiquid instruments, lack of **historical** data, etc.);

(Table Omitted)

Captioned as: Exhibit 2

(Graph Omitted)

Captioned as: Exhibit 4

(Table Omitted)

Captioned as: Exhibit 3

(Graph Omitted)

Captioned as: Exhibit 5

the degree of residual error (differences between the actual and the mapped **portfolio**, equivalents, etc.);

the impact of diversification (whether it magnifies, dampens, or does not affect differences across VAR calculations);

the impact of sampling issues (sufficiency of sample period, size, and breadth).

VAR, while an important advance in risk measurement, is only one aspect of an overall risk-management program. Different VAR methodologies and selection of the key decision factors for VAR are appropriate for different firms and depend upon many factors. These include the types of exposures, other qualitative and quantitative risk-management techniques employed, and the firm's risk appetite relative to its capital base. However, combined with the appropriate additional risk-management and risk measurement tools, VAR gets high marks.

Footnote:

Notes

Footnote:

1Mathematically, VAR quantifies the amount of expected loss based on the probability of certain market events occurring during a stated time period. 2A delta equivalent is a linear estimate of a security's **value** based on its first derivative with respect to a specific factor or factors.

Footnote:

3RiskMetrics is perhaps the most widely used of available data and assumes normal distributions.

4Fat-tailed behavior, also known as leptocurtosis, refers to distributions in which there is a broad range of values at the tails (for example, 1% of the time). 5Charles Smithson of CIBC Wood Gundy summed it up very well in a recent discussion regarding VAR: "The talk to action ratio is very high." 6Nonlinear price functions exist not only for options, derivatives with exponential functions, and leveraged instruments but also when yields are mapped into prices (for example, basic bonds). For an example of how these affect VAR, see Tanya Styblo Beder, "VAR: Seductive but Dangerous," **Financial** Analysts Journal, SeptemberOctober

1995.

Footnote:

7It is possible to run a simulation VAR that uses variance/covariance data, such as RiskMetrics. This technique is illustrated in the section, "VAR in Practice" (page 21).

8See, for example, Tanya Styblo Beder, "VAR: Seductive but Dangerous."

9See, for example, J. V. Jordan and R. J. Mackay, "Assessing **Value** at Risk for Equity Portfolios: Implementing Alternative Techniques," Handbook of Firmwide Risk Management, Beckstrom, Campbell, and Fabozzi, editors, forthcoming 1996, as reported in Risk Magazine, January 1996. Differences of more than 10 times are set forth in this data.

Footnote:

10See, for example, Darryll Hendricks, "Evaluation of **Value** -at-Risk Models Using **Historical** Data," FRBNY Economic Policy Review, April 1996. "See, for example, Philippe Jorion, "Risk2: Measuring the Risk in **Value** at Risk," **Financial** Analysts Journal, forthcoming

2Learning from these mistakes, firms often limit the portion of their **portfolio** or overall exposure that is based on theoretical mark-to-model values or erratic/infrequent data points. In addition, firms often impose the requirement that risk management, audit, IRO, or custodian obtain outside pricing from a different dealer than the dealer from whom the customized or illiquid securities were purchased.

Footnote:

The seminal work by Markowitz. 14See, for example, Tanya Styblo Beder, "VAR: Seductive but Dangerous."

15See, for example, Tanya Styblo Beder, "Derivatives: The Realities of Marking to Model," Bank Accounting & **Finance**, Summer 1994, 4.

16An assumption of anticipated or experienced market behavior.

17The market yield for each strip as of May 25, 1995, is stated on an actual/365 basis with semiannual compounding. The price of each strip is stated as a percentage of face amount.

Author Affiliation:

Tanya Styblo Beder is a principal, Capital Market Risk Advisors, Inc., New York. She thanks Frank Iacono for his valuable input regarding this article.

Special Features: Charts; Graphs; References
Descriptors: Financiallt; institutions; **Valuation**_ risk="risk" management="management" br="br">**Geographic Names:** US

36/9/11 (Item 11 from file: 268)
00299984
Moral hazard and Texas banking in the 1920s
Hooks, Linda M; Robinson, Kenneth J
Federal Reserve Bank of Dallas Financial Industry Studies , v 1 , n 96 , p 1-35 , Oct 1996 **Document Type:** Journal Article **Journal Code:** FIS **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 07426

Abstract:

Using recently collected examination data from a sample of Texas state-chartered banks over the period 1919-1926, the role of moral hard in increasing ex-ante asset risk is analyzed. During this period, a state-run deposit insurance system was in place that was mandatory for all state-chartered banks in Texas. Nationally chartered banks were not allowed to participate in the insurance program. An **analysis** of individual bank-level data reveals that declines in capitalization were positively correlated with increases in loan concentrations at insured banks. This is consistent with a moral-hazard effect at work. No such relationship is found between capitalization and risk at uninsured banks.

Text:

I. Introduction

Interest in the impact of deposit insurance on bank activities revived during the banking difficulties of the 1980s. A number of analysts argued that insured banks are subject to moral hazard. Under an imprecisely-priced deposit insurance program, banks face an economic

incentive to pursue additional risks than they would otherwise undertake in the absence of deposit guarantees. Thus, it is argued that one source of the **financial** difficulties of the 1980s was the existence of federal deposit insurance at banks and thrifts.

That deposit insurance may affect a bank's investment decisions has long been of concern to observers of banking. Opponents of government-sponsored deposit insurance in the early 1900s noted that insurance might serve as a subsidy from conservatively run banks to more aggressive banks. Despite these concerns, a number of states in the early 1900s adopted deposit insurance systems. In Texas, the state legislature of 1909 passed what was known as the Meachum-Greer bill (or Senate Bill Number 4), which established a deposit insurance program for all state-chartered banks.

Using data collected from bank examination reports from the 1920s, the experience of Texas banks under the state deposit insurance program is investigated in an effort to explore the role of insurance coverage in affecting the risk profile of banks. The Texas experience is especially interesting because the insurance program was mandatory for state banks, but national banks were not allowed to join. As such, nationally chartered banks can serve as a control group in the empirical tests of whether deposit insurance encourages banks to pursue more risky activities. To our knowledge, this is the first paper to examine the moral-hazard issue using individual bank data for a state insurance program in which participation was mandatory. Because we use newly-collected examination data on Texas banks beginning in 1919, we are able to provide a new data source for testing the view that deposit insurance has a significant impact on banks' investment decisions.

We find evidence consistent with moral-hazard incentives at work. Specifically, statechartered banks, which were covered by the state's deposit insurance system, pursued more risky activities following declines in their capital positions. We do not find such a relationship between capital and asset risk for uninsured national banks.

The paper proceeds as follows. Section II reviews the literature on moral hazard, followed by a description of Texas banking in the early part of the century in Section III. Section IV describes the deposit insurance programs implemented by the state of Texas, followed by a review of the new data on Texas banks in Section V. The empirical model is presented in

Section VI, followed by a discussion of the results in Section VII and our conclusions in Section VIII.

II. Previous Evidence on Deposit Insurance and Moral Hazard The idea that deposit insurance affects a bank's incentives is well established in economic theory. As Kareken and Wallace (1978) point out, a bank facing a fixed-rate deposit insurance system has an incentive to take additional risks, an example of the moral-hazard problem. Merton (1977) shows that by guaranteeing deposits, the insurer effectively has issued a put option on the assets of insured institutions. Deposit insurance gives the owners of an insured bank the right to sell the firm's assets to the insurer for the face **value** of deposits upon their effective maturity date, or the date of the next regulatory examination. The well-known comparative static properties of the standard options-pricing model highlight the perverse incentives inherent in fixed-rate deposit insurance systems -- the **value** of bank equity is maximized by increasing both leverage and asset risk (Smith, 1976 and Marcus, 1984). Moreover, the gains to increasing asset risk rise as a bank approaches insolvency.

Declines in capital then enhance the **value** of the put option inherent in deposit insurance and provide incentives for extra risk-taking. Kane (1989) identifies the "zombie" problem in which a **financial** institution with low capital has additional incentives to take on risk. As shown in Buser, Chen, and Kane (1981), and Marcus (1984), however, bank charter **value** can mitigate the risk-taking incentives of deposit insurance. If charter **value** is sufficiently large, **value** maximization can dictate protection of the charter, and in this case, increases in leverage and asset risk would reduce rather than enhance shareholder wealth.

Although moral hazard is well defined theoretically, only a few studies have attempted to estimate empirically the importance of this effect. Barth and Bradley (1989) find that the behavior of insolvent thrifts differed in some ways from their solvent counterparts. Beginning in 1982, mortgage assets, as a percent of total assets, declined at both solvent and insolvent institutions, but the decline was more pronounced at insolvent thrifts. Within the mortgage category itself, however, insolvent institutions rapidly increased their commercial mortgage lending relative to solvent thrifts, an area of lending generally associated with greater risk. Overall though, Barth and Bradley find little evidence that thrifts took significant advantage of new powers granted them in the early 19805.

Keeley (1990) offers empirical support to the view that increased competition and the concomitant decline in charter values exacerbated the moral-hazard problem among commercial banks. Gunther and Robinson (1991) find evidence of moral hazard in Texas banks during the 1980s. Brewer and Mondschean (1994) look at the extent to which underpriced deposit insurance has distorted the risk/return tradeoff for thrift shareholders. They find evidence that the stock prices of thrifts with a high risk of failure reacted favorably to shifts from traditional mortgage lending toward more volatile loans and investments, such as junk bonds.

While these studies generally find evidence consistent with moral-hazard incentives at work, it is not possible to know what choices banks would have made if they were not covered by deposit insurance. This issue arises in almost any study of present-day banking because almost all banks are covered by FDIC insurance. Thus, there is generally no control group without deposit insurance which could be used to distinguish the effects of deposit insurance from other influences on banks' decision-making.

To overcome this control problem, a number of studies have relied on **historical** data on banking when investigating moral hazard. Individual states experimented with deposit insurance many years before the federal government adopted it, and the experience of banks in these states provides insights into the effects of deposit insurance. Wheelock (1992), Wheelock and Kumbhakar (1995), and Wheelock and Wilson (1995), examine individual banks in Kansas that participated in a state deposit insurance

program in the 1920s, on a voluntary basis. These studies provide evidence that those banks which chose insurance coverage took additional risks.

Using data aggregated at the state level, Calomiris (1990a,1990b) finds that deposit insurance in the early 1900s increased incentives for bank risk taking, with the effect being stronger for states with mandatory insurance than for states with voluntary insurance. Looking at individual U.S. thrifts, Grossman (1992) determines that a thrift in the 1930s, when first admitted to the voluntary federal insurance program, did not take more risks than an uninsured thrift. But, after several years, risk did increase. In sum, these studies suggest that, based on individual bank and thrift data, moral-hazard incentives exist for voluntary insurance programs, and based on aggregated data, similar incentives exist for mandatory programs as well.'

To our knowledge, this paper is the first to use individual bank level data to examine an insurance program which was mandatory. Because Calomiris concludes from state-level data that mandatory deposit insurance enhanced the moral-hazard problem, the evidence for moral hazard should be stronger with the individual Texas bank data than for a state with a voluntary program. This paper builds on previous work estimating moral-hazard effects by incorporating proxies for ex-ante measures of risk. In particular, we are able to make use of two measures of loan concentration that have not been available in previous studies -- the proportion of loans secured by agriculture and the proportion secured by real estate. Finally, this paper provides additional evidence on moral hazard by examining individual bank behavior from a sample of nationally chartered banks that were ineligible to participate in the state's insurance system. The availability of these data, along with the structure of the Texas banking system in the early part of the century, offer a unique opportunity to examine whether the existence of deposit insurance heightened asset risk.

III. A Brief History of Texas Banking

Texas' attitude towards banking was one of ambivalence, if not outright hostility. At the state's constitutional convention, held in 1845, the following clause was adopted: "No corporate body shall hereafter be created, renewed, or extended, with banking or discounting privileges" (Grant and Crum, 1978, p. 15). Although nationally-chartered and private banks existed in Texas in the 1800s, the Texas legislature passed the Texas State Bank Law authorizing state-chartered banks only in 1905.2 A boom in state-chartered banking followed, and by 1910, there were a total of 584 state banks in Texas. Chart 1 shows that state banks quickly outgrew private banks in terms of asset holdings. However, nationally-chartered banks held the majority of the state's banking assets.

In part, the rapid increase in the number of state banks was the result of relatively modest requirements for establishing a state bank compared to a national bank. Obtaining a state charter required that the owners file an application with the Department of Insurance and Banking, but did not require further investigation by the state before the charter was granted. However, in

1913, the state legislature amended the law to require the state Commissioner of Banking and Insurance to investigate potential bank owners and the local market before approving a charter, thus effectively increasing the difficulty of obtaining a state charter.

The 1905 law also established a capital requirement for state-chartered banks which was lower than the capital requirement for nationally-chartered banks. As a result, the cost of opening a bank fell, and small towns that previously had few or no banks saw an increase in the number of state-chartered banks. In 1913, the Federal Reserve Act required each state bank that became a member of the Federal Reserve System to meet the capital requirements for a national bank, thus eradicating this advantage of a state charter for Federal Reserve member banks.

A state bank did have some flexibility in lending that a national bank did not have. A state bank could lend using real estate as collateral, as long as the bank did not lend more than 50 percent of its capital or more

than 50 percent of the market **value** of the collateral. A national bank could lend on real estate only up to an amount equal to 25 percent of its capital. Also, while a national bank could lend only up to 10 percent of its capital to a single borrower, a state bank could lend up to 25 percent of its capital to a single borrower.

Although the 1905 law encouraged the opening of new state banks, it also imposed a number of restrictions. One major regulation imposed was that shareholders of state banks faced double liability on their shares. Further, the law limited geographic expansion of a bank by allowing only "unit banking," which prohibited both state and national banks from establishing branches. Several years later, state-chartered banks were also required to join the newly created deposit insurance system.

IV. The Texas Deposit Insurance Program

Concerns about bank stability following the panic of 1907 led, in 1909, to the creation of a bank deposit insurance program organized by the Texas state government. The Depositors Guaranty Law mandated participation in the insurance program for all state-chartered banks in Texas. It also allowed national banks to participate on a voluntary basis. However, the Comptroller of the Currency had ruled in 1908 that national banks could not participate in state deposit insurance programs, so national banks in Texas did not join the state system. The insurance program became effective on January 1, 1910, and officially operated until February 11, 1927, although it experienced difficulties before 1927.4

In the Depositors Guaranty Fund,' a state bank paid an insurance premium into an insurance fund. The initial assessment for a member of the Guaranty Fund equaled one percent of the previous year's average demand deposits. In the following years, the annual insurance assessment was to be one-fourth of one percent of average demand deposits, or up to two percent in an emergency. A bank's assessment did not depend on the riskiness of its investments. The Guaranty Fund protected only non-interest-bearing deposits, but the amount of protection for these deposits was unlimited.

A state bank could also obtain insurance through a program known as the Bond Security System, in which a bank filed annually with the state a bond equal to the amount of its capital. The number of banks which chose this alternative was small, peaking in 1914 at 62 banks, or seven percent of all state banks (Grant and Crum, 1978).5 The proportion of banks in this alternative program remained small because a bank was not allowed to switch from one insurance program to the other until fairly late in the time period during which the Depositors Guaranty Fund operated.

As noted above, all state-chartere4 banks were required to participate in the insurance program, while no nationally-chartered banks were permitted to join. State bankers actually considered deposit insurance to be one of the main advantages of a state charter (Weaver, 1926).

Although few state banks wished to avoid the insurance program, a bank could do so if it was willing to convert to a national charter. In the first year of the program's existence, no state banks chose this option, and very few did so through the mid-1920s. This is likely the result of several factors, mainly the perceived advantages to marketing a bank as insured, and the relatively small cost of participating in the insurance program, which averaged only $50 annually in the 1910s (Weaver, 1926). Another important factor was that state law made charter conversion quite costly. According to the law, a state bank could convert to a national charter only if it liquidated completely and re-opened under a new national charter. In 1923, the legislature changed the law to allow a state bank to convert to a national charter without a complete liquidation. The number of conversions rose after this change, as noted below, but did not constitute a large portion of banks in any of the remaining years of the Depositors Guaranty Fund.

The Depositors Guaranty Law also mandated a relationship between bank size, in terms of deposits, and bank capital. In effect, the law established a capital-to-deposit ratio requirement, somewhat like current U.S. capital-to-asset requirements. A bank with $10,000 of capital

initially could acquire deposits equal to no more than five times its capital plus surplus. The law allowed progressively higher multiples of deposits relative to capital plus surplus, up to a maximum of ten times capital plus surplus for a bank with capital greater than $100,000.

The early years of the Depositors Guaranty Fund were uneventful. The 1910s were relatively stable years for banks, and thus for the insurance program.6 However, a downturn in economic activity in the early 1920s precipitated fairly widespread **financial** difficulties. Table 1 shows bank liquidations in Texas from 1919 through 1926. This time period coincides with the availability of our examination data and with the operations of the Depositors Guaranty Fund. The liquidation rate began to increase fairly steadily for state-chartered banks in the early 1920s before leveling off a bit. However, liquidations then jumped to almost 17 percent in 1925. Except for 1921, the liquidation rate of Texas national banks was below their state-chartered counterparts, sometimes by a fairly substantial margin. These differences in liquidation rates could reflect the impact of moral-hazard incentives at work at state-chartered banks.

The emerging banking difficulties of the 1920s began to strain the resources of the Depositors Guaranty Fund. Insurance assessments began to rise to meet increased demands on the insurance fund associated with the costs of covering insured deposits in failed banks. The Depositors Guaranty Fund began to charge special assessments on banks, over and above its regular assessments. These special assessments were imposed on banks in order to replenish the Depositors Guaranty Fund to its mandated $2 million. Table 2 shows that total assessments for all state banks climbed from less than one-tenth of one percent of total assets in 1919, to a peak of 1.68 percent of total assets in 1921. For the state banks in our sample, their assessment burden was initially less than the entire population of state banks until 1922. Insurance assessments peaked for our sample of state-chartered banks at 2.09 percent of assets in 1925.

As the cost of membership rose, some state banks decided to leave the program by converting their state charters to national charters, especially after 1923, when new state laws made conversion less costly than before. Thus began the erosion of the insurance program. Grant and Crum (1978, p. 49) report only a few conversions in 1924, and 80 conversions, or about ten percent of total state banks, in 1925.

The deposit insurance system deteriorated further in 1926, after the state legislature amended the law to allow a bank to move between the Depositors Guaranty Fund and the Bond Security System, discussed above. A mass exodus to the Bond Security System followed. This, of course, left the Depositors Guaranty Fund in an even more precarious situation, as it had fewer banks on which to levy assessments. A bank failure in September 1926 ultimately pushed the Depositors Guaranty Fund into insolvency. As a result, the state legislature repealed the Depositors Guaranty Law in February 1927, effectively ending Texas' experiment with deposit insurance.7

Despite its ultimate demise, the existence of a deposit insurance program for statechartered banks gives rise to the possibility that moral-hazard incentives contributed to the banking difficulties experienced in Texas during the 1920s. The recent collection of examination reports for a number of state-chartered banks in Texas over the period during which the Depositors Guaranty Fund was in operation provides us with an opportunity to add to the existing empirical literature on the role of moral hazard in increasing banks' risk.

V. New Data on Texas Banks 1919-1926

The empirical work uses newly collected data from Federal Reserve Bank of Dallas Examiner's Report of Condition. Along with basic balance sheet information, these examiners' reports include annual data on a bank's loan **portfolio** categorized by type of collateral used. A **random** sample of 89 banks out of approximately 200 available banks comprises our data set.8 These data represent an unbalanced panel data set because not all banks appear in each year. As Table 3 shows, the

sample observations per year capture one-quarter to almost one-half of all state member banks examined by the Federal Reserve in that year, with the sample banks' assets equal to between 10 and 20 percent of total state member bank assets by year.

While these new data provide an Interesting data base for a period in which little information on individual banks is available, they do have a number of limitations. First, the examinations were not regularly scheduled at the same time each year, so the time **interval** between panel observations is not constant. Second, the available archival records indicate that some examination reports have been destroyed, so the sample chosen from existing reports is not a sample drawn from the original population. We are also not able to determine which, if any, of these banks failed. In fact, we have been unable to determine why these examination records even survived. Thus, it is important to acknowledge that these factors might drive some of the empirical results reported.

To complement the data on state-Chartered banking, and to provide a control group, data were also collected from basic balance sheet information on nationally chartered banks in Texas. These banks were not covered under any deposit insurance system and would not be expected to exhibit any moral-hazard effects. This data set uses balance sheet information drawn from reports of condition from a **random** sample of 50 out of approximately 500 nationally chartered Texas banks that were in existence in 1919.9 The reporting dates of the national bank reports are the same in a given year for each bank, but at times differ slightly across years. As Table 4 shows, the national bank sample by year equals between 6 and 9 percent of the total number of national banks in Texas with between 10 and 22 percent of total national bank assets in the state.

Finally, Table 5 compares summary statistics for our sample of state and national banks to the population of state and national banks in Texas over the period 1919-1926, when the insurance program was in operation. The typical state-chartered bank in the sample recorded a capital-to-asset ratio of 19 percent, with over 60 percent of its assets in loans. These figures are slightly below those for the average of all state banks in Texas. The size of the average state bank in our sample equals almost $500,000 in total assets, slightly larger than the average of all state banks in Texas. On average, the national banks in our sample were better capitalized than all Texas national banks, but were less-well-capitalized compared to state banks. The loan-toasset ratio for our sample of national banks is identical to the Texas national bank population, and slightly below the comparable numbers for state banks. The average national bank in our sample recorded total assets equal to $2.5 million, and is thus significantly larger than the average state bank, and larger than the average size of all Texas national banks. From these comparisons, the sample observations used in our empirical estimations of moral-hazard effects appear to be reasonable representations of the populations from which they were drawn.

VI. The Empirical Model

Subjecting the theory of moral hazard to empirical testing is far from straightforward. Perhaps the biggest difficulty is quantifying shifts in banks' ex-ante risk-taking. While rapid asset growth is sometimes associated with higher risk, the structure of a bank's assets, measured by loan concentrations, can be a good proxy for risk. **Historically**, the loan-to-asset ratio has tended to move procyclically. Banks are more willing to extend credit during an upswing in economic activity and retrench their lending during downturns. A higher proportion of loans leaves a bank more exposed to credit risk and more vulnerable to an adverse economic shock. Using other measures of risk, such as the troubled asset ratio or the proportion of nonperforming loans in banks' portfolios, would seem less desirable. These are ex-post measures of risk that could reflect factors other than managerial choice, such as a downturn in economic activity. Moreover, if some banks chose to pursue more risky activities as capital declined, and these ventures proved successful, then they would not

be captured by ex-post measures of risk.

To estimate whether moral-hazard incentives were present, we test whether asset risk, as proxied by various loan concentrations, is related to prior capitalization levels among a sample of insured Texas banks operating during the 1920s. Consistent with the theoretical models of Merton (1977), Marcus (1984), and Keeley (1990), we hypothesize that those banks with declines in capital would be the ones most prone to exploit a deposit insurance subsidy and take on added risk. The concentration of assets in loans and the concentration of loans in selected categories serve as our proxies for ex-ante asset risk. If declines in the capital position of banks are associated with subsequent increases in loan concentration, then this would provide evidence consistent with a moral-hazard effect at work.

Chart 2 shows the distribution of average total loans over the time period 1919-1926 for our sample of state and national banks.'o While some state and national banks recorded fairly large loan concentrations, the state-chartered (insured) banks consistently exhibited larger loan to asset ratios. As stated above, a unique aspect of our data on state-chartered banks is the availability of loan concentrations based on the type of collateral used to secure the loan. Higher proportions of particular loan categories might be superior measures of ex-ante risk if they reflect a reluctance to diversify asset portfolios. This risk could presumably be reduced by altering **portfolio** compositions toward, say, greater holdings of government securities. Chart 3 shows the distribution of the average amount of both agricultural loans and loans secured by real estate for Texas state banks. Agricultural loans were much more important than real estate loans, and some of the banks in our sample exhibited, on average, fairly large loan exposures to this Particular type of credit."

But, while loan exposure was high at some of these banks, the moral-hazard effect would suggest that these loan concentrations might be negatively related to **financial** strength, as measured by capital. In an effort to determine if loan concentrations were negatively related to prior **financial** strength, we estimate the following model'2: where LOANASSj, is the loan to asset ratio for bank i at time t, and CAPASS;"., is the (lagged) capital to asset ratio for bank i. If a moral-hazard effect is related to prior capitalization levels, then the expected sign on P, is negative. The lagged loan concentration term is included to allow for a partial-adjustment process in lending activity. SIZE, is the log of a bank's assets, and is intended to control for bank size, which might be important given that the uninsured banks were larger institutions. To control for changes in economic activity, we use ECONOMY" which represents a vector of annual dummy variables, while Eil is an error term.13 Our data allow us to use several definitions for the dependent variable, including total loans (LOANASS), agricultural loans (AGLOANS), and loans secured by real estate (RELOANS). Before estimating the empirical model, we first offer some evidence on the efficacy of these proxies for ex-ante risk.

VII. The Results

VII. A. Loan Concentrations as Measures of Ex-Ante Risk

To be desirable proxies for ex-ante bank risk, loan concentrations should be positively related to the likelihood of failure. As stated above, however, we are unable to determine which, if any, of these banks failed. Instead, we examine the statistical significance of a bank's loan concentrations in explaining its subsequent troubled asset ratio (TROUBLED) by estimating the following equation:

If prior loan concentrations are statistically significant in explaining troubled assets, after accounting for size and economic activity, then this would support using loan concentrations as measures of ex-ante risk. The results from estimating equation (2) using the different measures of loan concentrations are found in Table 6. The coefficients on past values of LOANASS, AGLOANS, and RELOANS are all positive and statistically significant. To the extent that those banks with higher

proportions of troubled assets are more likely to fail, these results indicate that measures of loan concentration might be useful proxies for ex-ante risk."4

VII. B. Risk at Insured Banks

In an effort to detect if moral-hazard incentives increased banks' ex-ante risk, the empirical model given by equation (1) is estimated, with the results in Table 7. For our sample of Texas state-chartered banks, the model is estimated using first total loans as the dependent variable, and then using agricultural loans and real estate loans, all expressed as a percent of assets.15 The results in the second column of Table 5 use total loans as a percent of assets as the dependent variable and provide evidence consistent with a moral-hazard effect at work.'6 The coefficient on the capitalization variable is negative and significant, indicating that prior declines in capitalization give rise to greater asset risk, as measured by loan concentration. The coefficient on the lagged dependent variable and on bank size are also statistically significant.

The third column of Table 7 uses the ratio of agricultural loans to assets as the dependent variable. Again, results consistent with a moral-hazard effect are indicated by the negative and significant (at the ten-percent level) coefficient on the capitalization term. Finally, the results from using real estate loans are shown in the fourth column of Table 7. Here, the model does poorly. Only the lagged dependent variable is statistically significant. The poor results from using the concentration of real estate loans could reflect the relatively unimportant role real estate lending played for our sample of state-chartered banks, as revealed in Chart 3.

Overall, the results from these estimations provide some evidence that declines in capitalization led to increases in asset risk, as measured by loan concentrations."7 Such activity on the part of insured banks would indicate that banks with weakened **financial** conditions increased the riskiness of their asset portfolios, which would be consistent with a moral-hazard effect at work.'8

VII. C. Risk at Uninsured Banks

To provide further evidence on the role of moral hazard in increasing risk, we also estimated the empirical model using balance sheet data for nationally chartered banks in Texas over the same time period. These banks were not allowed to join the state insurance program, and thus their deposits were not covered by any guarantees. As such, these banks offer a control group with which to investigate the moral-hazard problem.19 Because this group of banks operated in some of the same market areas and over the same time periods as the state-chartered banks, the effects of economic activity should be similar between the two groups of banks20In the absence of a moral-hazard effect, we expect the coefficient on the capitalization variable to be either insignificant or positive. The results from estimating the empirical model for national banks are shown in the last column of Table 7. For this group of banks, we only have data on total loans. The capitalization variable is statistically insignificant, indicating that for this group of banks, ex-ante risk was not correlated with **financial** strength. Moreover, testing for differences in the coefficients in the model for uninsured national banks versus the comparable model for the insured state banks indicates that the coefficients on the capital term are statistically different from each other across the two models, as are the coeffficients on the lagged dependent variable and bank size. However, the coefficients on the annual dummy variables were not statistically different between these models, indicating that economic activity affected both groups of banks in a similar manner.21 These results provide some evidence that for uninsured banks, moral hazard was not present.22 When combined with the results using data for state-chartered banks, it appears that declines in **financial** condition gave rise to heightened asset risk. Such a response is consistent with moral-hazard incentives at work.

VIII. Conclusions

It has been well established theoretically that a fixed-rate deposit insurance system can provide banks with moral-hazard incentives to increase risk. Empirical verification of this potential for increases in risk at insured institutions has been difficult to obtain, however. In this paper, we use data on both insured and uninsured banks in Texas during the 1920s in an attempt to estimate whether a deterioration in **financial** condition at insured banks led to greater asset risk. Using different measures of loan concentration as our proxy for ex-ante risk, we find some evidence that a moral-hazard effect was at work for banks covered by a deposit insurance system. However, using data from a sample of banks operating in the same general area and same time period that were not covered by deposit guarantees, no significant relationship is indicated between loan concentrations and capitalization.

(Table Omitted)

Captioned as: Description of Variables

(Graph Omitted)

Captioned as: Chart 1

(Graph Omitted)

Captioned as: Chart 2

(Graph Omitted)

Captioned as: Chart 3

(Table Omitted)

Captioned as: Table 1

(Table Omitted)

Captioned as: Table 2

(Table Omitted)

Captioned as: TAble3

(Table Omitted)

Captioned as: Table 4

(Table Omitted)

Captioned as: Table 5

(Table Omitted)

Captioned as: Table 6

(Table Omitted)

Captioned as: Table 6

(Table Omitted)

Captioned as: Table 7

(Table Omitted)

Captioned as: Table 7

Footnote:

' On a related issue, White (1981) finds that states with unit banking were more likely to choose deposit insurance.

Footnote:

2 Grant and Crum (1978) provides a more complete history of the state-chartered banking industry in Texas, and this section draws upon it. For a comparison of state banks to national banks in Texas, see Shirley and Nichols (1931).

3 For state banks, the minimum capital needed ranged from $10,000 to $100,000, depending on the population of the city. For national banks, the requirement ranged from $25,000 to $200,000, again depending on population. In only one population category did the state requirement equal the national requirement; elsewhere, the state requirement was lower than the national bank requirement. See White (1983) for a summary of capital requirements by state.

Footnote:

4Robb (1921), Shibley (1914), and the FDIC Annual Report (1956) provide a comparison of the Texas law to other contemporary laws. Shibley includes a copy of the Texas law in an appendix.

5 None of the banks in our sample chose this form of deposit insurance.

Footnote:

6 Prior to the 1920s, the failure rate of Texas state and national

banks remained below one percent.

Footnote:

Over the period 1908-1917, seven other states -- Oklahoma, Nebraska, Minnesota, North Dakota, South Dakota, Kansas, and Washington -- also implemented various types of deposit insurance systems, all of which suffered the same fate as the Texas system. See American Bankers Association (1933), Calomiris (1989), and Wheelock (1992).

Footnote:

8 A **random** sample was necessitated due to the resource constraints encountered in collecting, tabulating, and entering the data from the examination records.

Footnote:

9 These data are published in Annual Report of the Comptroller of the Currency from 19191926.

Footnote:

'o While bank examination data are available through the early 1930s, we confine our empirical **analysis** to the time period 1919-1926 to coincide with the operation of the Depositors Guarantee Fund.

" Comparable data on loan concentration categories are not available for national banks. Some state-chartered banks also reported loans secured by stocks and bonds, but these loans represented relatively small components of total loans for these banks. 12 See the appendix for a complete description of the variables used.

Footnote:

13 We also try three explicit measures of economic activity in the regressions. First, we used annual cotton production quantities, in bales, by county. Cotton was the major crop in Texas during the sample period. According to The Texas Almanac (1931), Texas led the U.S. in cotton production each year in the sample, producing on average about one-third of the total U.S. crop. The annual **value** of the cotton crop in Texas during this period ranged between 44 and 63 percent of the **value** of all Texas crops, and approximately half of all Texas land devoted to crops was planted in cotton. Second, we used the annual number of building permits issued in five regions of the state, and then the dollar **value** of permits issued. These measures were intended to capture some of the non-agricultural activity which cotton production misses. Cotton data were obtained from The Texas Almanac (1933), and data on permits and their **value** were obtained from various issues of the Federal Reserve Bank of Dallas' Monthly Review of Business and Industrial Conditions, 1920-1927. However, these variables were not statistically significant in any of the models estimated.

Footnote:

14 The coefficients on the annual dummy variables are not shown, but they were jointly significant. In addition, because the data on state bank examinations are for different times during the year, and because agricultural loans are such an important component of the state banks' asset structure (see Chart 3), the models are also estimated using quarterly dummy variables in an attempt to capture the seasonal nature of agricultural lending activity. These quarterly dummy variables, while not shown, were also statistically significant in all of the models estimated.

Footnote:

15Two possible sources of endogeneity in this model are choice of charter and bank capital levels. As noted in Section IV, state banks converted to national charters very infrequently during our sample period because it was very costly. Therefore, selectivity bias would appear to be small. For capital, the model attempts to control for endogeneity by using the previous period's capital level. Such a predetermined variable would not be expected to show endogeneity. Thus, we believe endogeneity is unlikely in our model. 16Again, annual dummy variables are used to account for economic activity and quarterly dummy variables to account for the seasonal nature of agricultural lending. While not shown, these dummy variables are statistically significant in all of the models estimated.

Footnote:

17 To judge the robustness of the results, several alternate specifications were also estimated. First, the examination reports for the state-chartered banks indicate that, for some of the banks, the exams took place during different time periods over the years. To judge what effect this unequal spacing of observations might have on the results, for each bank we created a variable called MONTHS, which measures the number of months between examinations. We then interacted MONTHS with the lagged variables in our model, CAPASS,_1 and LOANASSt-,, in an attempt to judge whether these estimates are affected by irregular exam dates. These interaction terms were statistically insignificant, providing some evidence that the irregular nature of the exam data did not affect our results. We also used MONTHS interacted with the lagged loan concentrations when estimating equation (1). Again, these interaction terms were not statistically significant.

Footnote:

18 We also estimated the model with an interaction term defined as CAPASSt,*LOANASS,, which is intended to capture the possibility of both liquidity and regulatory influences on bank lending activity. These factors might be expected to mitigate the effect of changes in capitalization on lending activity. Those banks that already have a high amount of loan exposure might only be able to increase lending further only when capitalization increases due either to liquidity constraints or regulatory actions. The expected sign on the coefficient of this interaction term is thus positive. However, in the models estimated, this interaction term was statistically insignificant. We also estimated the various models using only those banks with five or more examination records to judge whether the frequency of examinations might affect the results. The same qualitative results found in Table 7 were obtained when limiting the sample to those banks with five or more examination records.

Footnote:

"9 Moreover, because the state insurance program was mandatory, and switching charters was relatively expensive, the adverse selection problem discussed in Wheelock (1992) and Grossman (1992) should not be very important.

Footnote:

20 The state banks in our sample operated in 53 different counties across Texas. The national banks in our sample also operated in about one-fourth of these counties. National banks tended to operate in larger counties. Using data from the 1920 census, the average market size (judged by county population) was 38,009 for national banks and 25,530 for state banks. The median values were 25,456 and 18,080, respectively. Therefore, in an attempt to control for possible differences in market areas due to location, we also estimated equation (1) using both the cotton and permits variables discussed in note 13 above. These measures are intended to capture differences between urban and rural areas. However, neither of these variables was statistically significant, indicating that the differences in market areas between state and national banks were not too large.

Footnote:

2 In testing the differences in the coefficients on CAPASS, across the equations for national and state banks, the **value**lt; of the t-statistic is 4.6. For differences between the coefficients on LOANASS,-, the t-statistic is 3.9, and on SIZE,, the t-statistic is 5.2.

Footnote:

22 One potential shortcoming of using loan concentration as a measure of ex-ante risk is the possibility that banks, when facing **financial**lt; difficulties, could elect to sell off some of their security holdings. In such a scenario, banks would see an increase in their loan to asset ratio following a decline in their capitalization, but this would be unrelated to increases in risk. To investigate this possibility, we regressed banks' securities holdings (as a percent of assets) on their

lagged capital to asset ratio (as well as size and the dummy variables). No statistically significant relationship was found between banks' holdings of securities and prior levels of capitalization, indicating that this type of asset restructuring was not behind the empirical results.

Reference:

References

Reference:

American Bankers Association (1933), The Guarantee of Bank Deposits Barth, James R. and Michael G. Bradley (1989), "Thrift Deregulation and Federal Deposit

Insurance, "Journal of **Financial**lt; Services Research 2 (September): 231-259. Board of Governors of the Federal Reserve System (1959), All-Bank Statistics, United States 1896-1955, Washington, D.C.: Board of Governors of the Federal Reserve System.

Reference:

Brewer, Elijah III and Thomas H. Mondschean (1994), "An Empirical Test of the Incentive Effects of Deposit Insurance: The Cm of Junk Bonds at Savings and Loan Associations," Journal of Money, Credit, and Banking 26, No. I (February): 146-164. Buser, Stephen A., Andrew H. Chen, and Edward J. Kane (1981), "Federal Deposit Insurance, Regulatory Policy, and Optimal Bank Capital," The Journal of **Finance** 35 (March): 5160.

Calomiris, Charles W. (1989), "Deposit Insurance: Lessons from the Record," Economic Perspectives, Federal Reserve Bank of Chicago 13 (May/June): 10-30.

Reference:

Calomiris, Charles W. (1990a), "Is Deposit Insurance Necessary? A **Historical** Perspective," The Journal of Economic History 50, No. 2, (June): 283-295. Calomiris, Charles W., (1990b), "Do Vulnerable Economies Need Deposit Insurance? Lessons from the U.S. Agricultural Boom and Bust of the 1920s," in Philip Brock, ed., If Texas Were Chile: **Financial** Risk and Regulation in Commodity-Eiporting Economies, (Washington, DC).

Reference:

Comptroller of the Currency, (1919-27), Annual Report, Washington, D.C. Federal Deposit Insurance Corporation (1956), Annual Report, Washington, D.C.

Reference:

Federal Reserve Bank of Dallas (1919-29), Annual Report, Dallas, Texas. Federal Reserve Bank of Dallas (1919-1926), Examiner's Report of Condition, Department of Examination.

Reference:

Federal Reserve Bank of Dallas (1920-1927), Monthly Review of Business and Industrial Conditions, Dallas, Texas.

Grant, Joseph M. And Lawrence L. Crum (1978), The Development of State-Chartered Banking in Texas from Predecessor Systems Until 1970, Austin, Texas: The University of Texas Bureau of Business Research.

Grossman, Richard S., (1992) "Deposit Insurance, Regulation, and Moral Hazard in the Thrift Industry: Evidence from the 1930s," American Economic Review, 82:4, 800-821.

Reference:

Gunther, Jeffery W., and Kenneth J. Robinson, (1991), "Empirically Assessing the Role of Moral Hazard in Increasing the Risk Exposure at Texas Banks," Federal Reserve Bank of Chicago, Proceedings of the 27th Annual Conference on Bank Structure and Competition, (May): 199-220.

Kane, Edward J. (1989), The S&L Insurance Mess: How Did It Happen,? Washington, D.C:

The Urban Institute.

Kareken, John H. And Neil Wallace (1978), "Deposit Insurance and Bank Regulation: A PartialEquilibrium Exposition," Journal of Business 51 (July): 413-38.

Reference:

Keeley, Michael C. (1990), "Deposit Insurance, Risk, and Market

Power," American Economic Review 80 (December): 1183-1200.
Marcus, Alan J. (1984), "Deregulation and Bank **Financial** Policy," Journal of Banking and **Finance**, 8: 557-565.
Reference:
Merton, Robert C. (1977), "An Analytical Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory," Journal of Banking and **Finance**: 3-11.
Robb, Thomas Bruce, (1921), The Guarantee of Bank Deposits, Boston: Houghton Mifflin Co. Shibley, George H., E1914), History of Guarantee of Bank Deposits in the States of Oklahoma, Texas, Kansas, Nebraska, and South Dakota 1908-1914, U.S. Senate, 63rd Congress, Second Session. U.S. Senate Document No. 522. June 23.
Reference:
Shirley, R.V., and Bervard Nichols (1931), Trends of Development of Texas **Financial** Institutions, Austin, Texas: University of Texas Bureau of Business Research. Smith, Clifford W., Jr. (1976), "Option Pricing: A Review," Journal of **Financial** Economics 3
(January/March), 3-51.
The Texas Almanac and State Industrial Guide (1931), Dallas, Texas: A.H. Belo Corporation. Weaver, Findley (1926), "State Banking in Texas," master's thesis, University of Texas, Austin. Wheelock, David C. (1992), "Deposit Insurance and Bank Failures: New Evidence from the 1920s," Economic Inquiry 30:3 (July), 530-43.
Reference:
Wheelock, David C. and Subal C. Kumbhakar (1995), "Which Banks Choose Deposit Insurance? Evidence of Adverse Selection and Moral Hazard in a Voluntary Insurance System," Journal of Money, Credit, and Banking, 27:1 (February),186-201.
Wheelock, David C. and Paul W. Wilson (1995), "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency," Review of Economics and Statistics, 689-700. White, Eugene N. (1981), "State Sponsored Insurance of Bank Deposits in the United States, 1907-1929," The Journal of Economic History 41 (September), 537-57
Reference:
White, Eugene N. (1983), The Regulation and Reform of the American Banking System, 19001929, Princeton: Princeton University Press.
White, Halbert (1980), "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica 48 (May): 817-38.
Appendix:
Appendix

Special Features: Charts; Graphs; Appendix; Equations; References
Classification: 9190 (CN=United States); 8100 (CN=Financial services industry); 9130 (CN=Experimental/Theoretical)
Descriptors: Moral hazard; Safety & soundness; Studies; Asset management; Risk management; State banks; Deposit insurance
Geographic Names: Texas

36/9/12 (Item 12 from file: 268)
00289259
Guarding against risk

Pierzchalski, Larry
Mortgage Banking , v 56 , n 9 , p 38-45 , Jun 1996
CODEN: MOBAAX **Document Type:** Journal Article **ISSN:** 0730-0212 **Journal Code:** MOB
Language: English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 04557

Abstract:

The use of credit scores from the Fair, Isaac and Co., Inc. (FICO) score model by mortgage lenders skyrocketed following the release of an industry letter by Freddie Mac in July 1995 stressing the benefits of credit scoring. Mortgage Guaranty Insurance Corp. (MGIC) in February 1996 unveiled its mortgage scoring model, which represents yet another advance in the mortgage industry's quest for even greater precision in mortgage risk assessment. The quest has intensified in recent years as the mortgage industry has come under pressure to increase loan volumes and make inherently riskier loans without increasing credit losses. Analyses show that the MGIC model is more predictive than the FICO credit score when the probability being measured is foreclosure within 4 years of origination. The FICO credit score alone is highly predictive and is, in fact, a key component of the MGIC model. Often, automated underwriting and scoring systems will flag a loan's strengths and weaknesses, drawing the attention of the underwriter to the portion of the loan that needs the greatest review. This focuses the underwriters' efforts, making them more efficient.

Text:

Headnote: The evolution of risk management in the mortgage industry is leading to a more precise understanding of the risk factors that affect loan performance. This path is leading the industry to mortgage scoring.

LAST JULY, FREDDIE MAC RELEASED AN INDUSTRY LETTER THAT catalyzed the mortgage industry's acceptance and use of credit scoring. The concept of giving a borrower a score that represents his or her probability of defaulting on a loan was new to many in the mortgage industry. Credit scoring has been used in consumer lending for four decades, but only a handful of companies in the mortgage industry were using it. * Then came the letter. 4 "After reviewing a number of alternatives, we determined that, within the manual underwriting process, one of the easiest and most readily available tools to assist you in managing the challenging credit-risk environment is the use of [credit scores]," wrote Michael K. Stamper, Freddie Mac's executive vice president for risk management. 4 The use of credit scores from the California-based Fair, Isaac and Co., Inc., (FICO) scoring model skyrocketed. Many lenders now realize that credit scores are useful in quantifying at least one of the "Three Cs" of mortgage lending-collateral, capacity, and credit. But what about the rest? Will there someday be a score that quantifies the risk of an entire mortgage? The answer is "yes." As the mortgage industry continues to adapt to credit scoring, several mortgage industry players are hard at work developing, or in some cases already marketing and using, mortgage scoring models designed to predict mortgage performance using all pertinent facts about the loan, the borrower and the collateral. Among them are major mortgage insurance companies, some large mortgage lenders and even the secondary market corporations.

Mortgage Guaranty Insurance Corporation (MGIC), Milwaukee, Wisconsin,

in February unveiled its mortgage scoring model, the MGIC Loan Performance Score. The scoring model represents yet another advance in the mortgage industry's quest for even greater precision in mortgage risk assessment. It is a quest that has intensified in recent years as the mortgage industry has come under pressure to increase loan volumes and make inherently riskier loans without increasing credit losses.

The belief is that the more precise the industry is in understanding mortgage risk, the more capable it will be of discerning applicants who are most likely to maintain long-term homeownership from those who are likely to suffer the pain of foreclosure. Some also believe that mortgage scoring will result in more objective and consistent underwriting, quicker decisions, increased productivity, lower costs and a simpler process for both borrower and lender. This, in turn, should lead to increases in loan volume while keeping loan losses at reasonable levels. That's an equation for greater profitability.

This is the vision. It is a vision that is driving industry leaders toward a greater understanding of mortgage risk; and mortgage scoring is being hailed as a risk-management tool that can help the industry put its knowledge of risk to work in underwriting, processing, quality control and loan **portfolio analysis**.

Support for credit scoring

(Graph Omitted)

(Graph Omitted)

Credit scores have been touted as a very valuable decisionsupport tool by many mortgage industry leaders. In its July 1995 industry letter titled "The Predictive Power of Selected Credit Scores," Freddie Mac encouraged lenders to obtain credit scores for each mortgage applicant and use them "as a tool" to focus underwriting and quality-control efforts.

Freddie Mac supported its suggestion by citing an **analysis** it had done that shows FICO credit scores are highly predictive of mortgage default (see Figure 1). The agency says it studied the performance of hundreds of thousands of loans originated over several years from a wide distribution of lenders, loan types and geographic markets. The credit scores the agency used in its extensive statistical **analysis** were obtained at or near the time of origination. Freddie Mac believes its findings are so conclusive that in the letter it identified two key credit score thresholds and suggested that mortgage lenders vary their depth of underwriting review on the basis of where a borrower's credit score falls on the FICO scale of 360 to 900 in relation to these thresholds. For example, applications from individuals with credit scores of less than 620 should receive a "cautious" review, Freddie Mac said. A comprehensive review was suggested for applications from individuals with scores of 620 through 660. And for those with scores greater than 660, Freddie Mac suggested a basic review. (The higher the score, the lower the risk.) The industry letter was unprecedented-no agency had ever before recommended the use of credit scores.

In October 1995, Fannie Mae issued its own letter of endorsement. "As a result of our own research into the use of credit scores, we have concluded that they can be an effective tool for evaluating credit risk because they result in increased accuracy, objectivity and consistency in the mortgage origination and underwriting process," the agency stated in its letter to lenders.

MGIC conducted its own **analysis** that corroborates the findings of Fannie Mae and Freddie Mac (see Figure 2). Our most recent research shows a strong correlation between the FICO credit score and the incidence of default within the first year of origination-referred to as early payment defaults (EPDs). In our **analysis**, borrowers with credit scores of 700 or greater were used as the base for comparison. We found that compared with the baseline, borrowers with scores of less than 620 were nearly 20 times more likely to miss monthly payments within the first 12 months of origination.

Borrowers with "thin" credit profiles-a credit history too short for a

score to be calculated-were 14 times more likely to default within a year from origination than borrowers with credit scores of 700 or higher.

MGIC has been using credit scores for about 18 months in underwriting and for two-and-a-half years in quality control and risk-management modeling. We use credit scores because we believe they are an objective and consistent measure of risk, as well as a very predictive risk-assessment tool.

However, credit scores alone do not implicitly incorporate the layering of risk that mortgage asset risk management has taught us can have a profound impact on mortgage performance. The fact that a score created primarily for consumer lending can be so predictive of mortgage default indicates just how important borrower credit is in dictating mortgage performance. But other factors are also important. That's why several mortgage industry leaders are pointing to mortgage scoring as a decisionsupport tool that will more precisely quantify mortgage risk.

Defining mortgage scoring

Consumer credit scores and mortgage scores are designed to measure different probabilities. Consumer credit scores measure a borrower's probability of defaulting on a consumer loan, whereas mortgage scores measure the probability of a mortgage going into default or foreclosure.

Because of this difference, credit and mortgage scoring models consider different variables. Consumer credit scoring considers essentially the same variables as those that underwriters review when assessing a borrower's credit profile.

FICO credit scores are derived from the consumer credit data kept on file at the nation's three largest credit repositories-Equifax, TRW and TransUnion. According to FICO, its scoring model groups the various credit variables obtained from these repositories into the following five categories: previous credit performance, a borrower's current level of indebtedness, the amount of time a borrower has been using credit, a borrower's pursuit of new credit and the types of credit a borrower has established.

In mortgage scoring, the variables are basically the same as those an underwriter would review before rendering a mortgage lending decision. Mortgage scoring systems typically consider four types of information-borrower **financial** and credit characteristics, loan product information, property **valuation** and local-market economic conditions. The credit score, loan-to-**value** (LTV) ratio, type of loan (refinance or purchase), type of loan instrument (adjustable rate or fixed rate), property type, debt ratios and health of the local real estate market are just a few of the individual risk factors typically considered in mortgage scoring.

Most systems designed to predict mortgage performance start with credit scores-a testament to their predictiveness. But the inclusion of other variables particular to mortgage lending greatly enhances the ability of these systems to predict mortgage performance.

Another key difference between consumer credit and mortgage scores is that credit scores are borrower-specific in that they represent the general likelihood of a particular borrower going into default on any loan, whereas mortgage scores are loan-specific in that they represent the likelihood of a particular mortgage going into default or foreclosure.

Thus, mortgage scoring shifts the focus of underwriting from a borrower's credit profile to the whole loan package. This means that mortgage scores can identify a mortgage as being lower risk, even when a borrower's credit score dictates that he or she poses a higher risk of default.

There are people in the mortgage industry who will ask: "How can this be?" With the release of Freddie Mac's findings a year ago, many in the mortgage industry have come to the strict conclusion that an individual with a 620 credit score poses a higher risk of poor mortgage performance than someone with a 660 credit score. While this is true most of the time, there are exceptions. Mortgage scoring identifies those exceptions, thanks

to its broader understanding of which risk factors influence mortgage performance.

The evolution of risk management

Mortgage scoring is the product of two occurrences: the evolution of traditional mortgage risk management and the application of modeling technologies in the mortgage industry.

In the mortgage insurance industry, risk management is the practice of studying **historical** loan performance to isolate the factors that contribute to default and foreclosure. Given this definition, there is no question risk management has evolved immensely in the last two decades. As one mortgage industry executive recently quipped: "Twenty years ago we laid our hands on a loan and blessed it; today, we actually underwrite it."

Sound underwriting execution is the key to maintaining overall loan quality; and risk management is the key to providing the proper direction to underwriters-the industry's stewards of quality. Today, not only do we, as an industry, underwrite, but we make a concerted effort to consider all the risk factors of a loan in combination, rather than any one factor in isolation.

In the last 10 years, our understanding of how certain risk factors affect loan performance, both alone and in combination with others, has grown by leaps and bounds. This has helped the mortgage industry provide more precise direction to underwriters in terms of understanding compensating factors and identifying appropriate offsets to high-risk factors.

It also has allowed the industry to liberalize certain underwriting guidelines, such as loan-to-**value** (LTV) ratios, debt ratios and the amount of reserves required after closing. The result is that underwriters today are directed less and less by intuition and more and more by empirical analyses of loanperformance data.

This industrywide push to find out what leads to foreclosure grew in the mid-1980s as the mortgage industry was besieged by a rising tide of defaults. The arrival of new, untested variable-rate mortgage products, coupled with a downturn in certain real estate markets, delivered a double whammy to the mortgage industry. In the midst of its woes, the industry's quest for more precise understanding of mortgage risk led to more in-depth **analysis** of **historical** loan-performance data.

In 1987, MGIC released a report titled "The Changing Nature of Risk in Mortgage **Finance**" that put the mortgage risk trends of the 1980s into perspective. "Though the risks associated with the modern lending environment are clearly greater and more complex," the report noted, "they can be effectively managed if they are fully understood."

MGIC was not alone in the quest for understanding. By the late 1980s, the mortgage industry focused on understanding product- and market-related risks and the impact they had on a borrower's ability and willingness to repay. Terms such as "payment shock" and "equity erosion" became common. The understanding of how employment trends affected housing occupancy and prices grew; and the emphasis on improved quality control intensified as Freddie Mac, at one point, announced mandatory buyback of any mortgage that defaulted within four months of origination. Economic and business conditions improved by 1990 and the quest for precision was driven less by an immediate need to minimize losses and more by the desire for more efficient underwriting and processing, and for greater flexibility to liberalize underwriting guidelines. The goal of expanding underwriting guidelines was to increase loan volume without compromising loan quality; and the industry's growing knowledge of mortgage risk played a significant role in allowing agencies, lenders and insurers to do this comfortably.

In a classic "half-empty, half-full" analogy, risk management can be viewed as the practice of either identifying risks to avoid or identifying risks worth taking. Those who take the "half-full," or more optimistic, view believe the quest for precision has become inextricably linked with efforts to expand homeownership in America.

The layering of risk

(Graph Omitted)

(Graph Omitted)

If risk management's evolution is the legacy of the 1980s, the understanding of risk layering is the legacy of the first half of the 1990s. It also is vital to building an effective mortgage scoring model.

In 1995, MGIC shared with the industry its **analysis** of the performance of loans originated from 1985 through 1990. While the **analysis** put in perspective the impact certain risk factors have on loan performance, more importantly it crystallized the concept of risk layering by showing how risk factors, when piled one on top of another, can dramatically increase the riskiness of a mortgage.

Our **analysis** looked at MGIC-insured mortgages that had at least five years of seasoning, or enough time to run the claim cycle. We focused on loans made in the Midwest to minimize the impact of "peak-and-valley" home price conditions experienced during that period by more volatile housing markets in the West and Northeast. The purpose of the **analysis** was to determine the magnitude to which certain risk factors, considered both individually and in concert with others, influence mortgage performance.

A year ago in this publication (June 1995, Mortgage Banking, "Ready to Make the Grade"), MGIC Executive Vice President Gordon H. Steinbach reported the results of this **analysis**. We have since updated our **analysis** to include mortgages originated from 1986 through 1991.

We studied the impact of risk layering by starting with a group of loans that had minimal high-risk factors (see Figure 3). The claim incidence of this group of loans was set at 1.0 on the relative claim index, and the performance of all other groups of loans was measured against this baseline pool of mortgages. The baseline grouping included 90 percent LTV mortgages to borrowers with housing debt-to-income ratios below 33 percent, total debt-to-income ratios below 38 percent, and an established credit history with no adverse credit items.

The first high-risk factor we examined was LTV. By increasing the LTV from 90 to 95 percent, while keeping all other factors the same, the relative claim rate doubled to 2.0. Then, while keeping the LTV at 95 percent, we pushed debt ratios higher in excess of 33 and 38 percent. The result was a relative claim rate of 2.9, or nearly triple the baseline. The last risk factor we added to the mix was adverse credit. The result was a nearly ninefold increase in the claim index to 8.8.

To many practiced industry observers, these results were very intuitive; and, in fact, many underwriters felt the **analysis** did nothing more than quantify the correlation between certain risk factors and mortgage performance they believed always existed.

MGIC has since directed its underwriting staff to be keenly aware of the excess layering of risk, while at the same time adopting its affordable housing underwriting guidelines as standard guidelines. Clearly, the evolution of risk management has led to a greater understanding of the layering of risk. Now the challenge facing the mortgage industry is to incorporate that understanding into the underwriting process with the help of technology.

Meeting the challenge

Effective mortgage-scoring models will help the industry meet this challenge. To understand how mortgage scoring will help, it is important to understand first how mortgage scoring is developed.

Basically, three ingredients are needed to develop an effective scoring model. The first is a data base of mortgage loans that has both the breadth and depth necessary to develop a model that accurately predicts what the developers set out to predict.

Also, the development data is vital in determining how broadly applicable a model is. For example, if only loans to borrowers with credit scores greater than 660 are used, then the resulting model likely won't be able to consistently weigh the impact lower credit scores have on probable mortgage performance.

The second ingredient is an understanding of mortgage risk. The concept of risk layering shows how far the industry has come in this regard. The model's developers must determine which risk factors are considered and how they are considered. In other words, they decide to test for the correlations between, say, debt ratios, credit history and loan instrument. Or they might decide to drop a variable from the mix because it is impeding or have no impact on the model's predictiveness.

The third ingredient in developing an effective scoring model is modeling methodology. The two types of statistical methodology typically used in the development of scoring models are regression **analysis** and neural net, a multidimensional, nonlinear approach to **analysis** that requires highpowered technology.

Neural-net modeling technology identifies the most predictive variables and analyzes them, both individually and in concert with other variables, to set the optimal weighting. Our testing has shown that neural-net technology develops a more predictive model than other approaches, such as regression **analysis**.

Neural-net technology is best known for its use by the U.S. Department of Defense in developing cruise missile guidance systems. Credit card companies currently use it to detect potential incidents of fraud. The neural net's strength is its ability to recognize aberrations in patterns of credit card use to help stop the use of stolen or bogus cards.

The MGIC mortgage scoring model was developed over the course of about three years using neural-net modeling technology. Loan-performance data for about 560,000 mortgages originated from 1989 through 1992 and insured by MGIC was analyzed by MGIC's Risk Management Department. From those 560,000 loans, MGIC's risk analysts **randomly** selected a subset of loans that they used to "train" the model.

The analysts then started with a set of about 40 variables they thought to be predictive of mortgage performance based on past statistical analyses done by MGIC. The analysts used the neural net to provide direction as to what variables were most predictive in a process that involved running the data through the neural-net software, analyzing it using traditional statistical methods, and then going back to the software for more discovery.

Eventually, the factors that **historically** have had the greatest impact on mortgage performance were not only identified, but were weighted against one another in an intricate matrix that forms the heart of the MGIC model. In all, 12 factors were identified as having the greatest impact on mortgage performance. The weighting of those factors implicitly incorporates the impact of risk layering. The result is a model that produces scores ranging from zero to 1,000, with higher scores indicating less risk.

Even though MGIC unveiled its mortgage-scoring model in February, the risk analysts are continuing their research and **analysis** of **historical** loan performance in the quest for even greater precision.

How does it compare?

In February, when the MGIC Loan Performance Score was unveiled, lenders wanted to know: "How does it compare to the FICO credit score?"

The analyses we've done show the MGIC model to be more predictive than the FICO credit score when the probability being measured is foreclosure within four years of origination. This should not be a surprise; nor should it be misconstrued as an indictment of the FICO credit score. The FICO credit score alone is highly predictive and is in fact a key component of the MGIC model.

The goal of our modeling process was to maximize the model's ability to differentiate between mortgages that will perform and mortgages that won't perform. If a scoring model's ability to differentiate were perfect, for instance, it would classify all mortgages that had gone into foreclosure in its highest-risk score category and all performing mortgages in its lowest-risk score category.

A model with minimal ability to differentiate might classify only a few more foreclosures in its highest-risk score category than in its lowest-risk score category. Deploying such a model would be extremely detrimental, if not catastrophic, since it would result in increased loan losses and denial of mortgages to applicants actually qualified to become homeowners.

One **analysis** MGIC conducted was a comparison of the scoring models' ability to differentiate high-risk from low-risk mortgages. For purposes of the **analysis**, we used the 560,000 loans originated from 1989 through 1992. We made two listsone ranking those 560,000 loans in order of their FICO credit score, the other ranking them in order of their MGIC score. We then cut those lists into five groups of 112,000 loans, each group representing 20 percent of our entire sample.

One measure of differentiation is to compare the highest-risk 20 percent with the lowest-risk 20 percent and simply quantify how many more foreclosures appeared in the grouping representing the riskiest loans. By this measure, the highest-risk 20 percent on the FICO list included eight times more foreclosures than its lowest-risk 20 percent. By comparison, the highest-risk 20 percent on the MGIC list included 34 times more foreclosures than its lowest-risk 20 percent (see Figure 4). In other words, the MGIC model's ability to classify performing loans in its lowest-risk quintile and foreclosures in its highest-risk quintile was more than four times greater than the ability of the FICO model.

The results of another comparative **analysis** MGIC conducted shows how borrowers with credit scores of less than 620 can in some cases actually be better risks than borrowers with credit scores of greater than 660. Figure 5 shows relative foreclosure rates within seven MGIC Loan Performance Score groupings. For purposes of this **analysis**, we again scored the 560,000 loans. But instead of dividing them into five equal bins of 112,000 loans each, we simply categorized them by placing them in one of seven score ranges based on their MGIC Loan Performance Score.

The baseline of 1.0 on the y axis represents the average foreclosure rate for the entire pool of 560,000 mortgages. On the x axis, the relative foreclosure rate is 1.0 for the 450-to-549 score range. We defined that range as the midpoint for purposes of seeing whether mortgages in the other score ranges were likely to perform above or below the average foreclosure rate of the entire pool of 560,000 mortgages.

Loans in the score ranges falling to the right of the midpoint range, for example, could be said to have a higher probability of outperforming the pool (i.e., a foreclosure rate lower than the pool's average foreclosure rate); whereas loans in score ranges falling to the left of the midpoint could be said to have a higher probability of underperforming the pool (i.e., a foreclosure rate higher than the pool's average foreclosure rate).

The ability of the mortgage score to implicitly incorporate risk layering and a precise understanding of mortgage risk is evident in the distribution of mortgages across those seven mortgage score ranges at various FICO credit-score **intervals**. When we looked at the subset of those 560,000 mortgages in which borrowers had FICO scores of less than 620, we noticed that about 22 percent of these mortgages scored at or above the midpoint range (see Figure 6). In the subset of loans in which borrowers had credit scores of 620 through 660, about two-thirds of the loans fell at or above the midpoint range (see Figure 7). And in the subset of loans in which borrowers had credit scores of greater than 660, about 90 percent fell at or above the midpoint range (see Figure 8).

(Graph Omitted)

(Graph Omitted)

So while it is true that an individual with a 620 credit score typically poses a higher risk of poor mortgage performance than someone with a 660 credit score, there are exceptions. Mortgage scoring, bolstered by the inclusion of variables in addition to borrower credit, helps

identify those exceptions.

The scoring steamroller

When Freddie Mac issued its industry letter touting the predictive power of credit scores last July, many industry leaders knew what would follow. They knew a public debate over credit scoring and its fair lending implications would ensue; that underwriters and processors would get an uneasy feeling envisioning a day when they are replaced by a scoring model; and that, despite these issues, mortgage lenders would flock to scoring as a way to bring greater efficiency to underwriting and processing and improve credit decisions.

(Graph Omitted)

(Graph Omitted)

Because of these issues, however, the industry needs to take a studied and cautious approach to implementing scoring. Companies looking to reap the benefits of improved efficiency and loan quality should step back and review how scoring can best be deployed to improve underwriting and processing. Lenders with automated decision technologies already in place will need to assess the technical adaptability of scoring to their systems.

Management also will need to provide the proper direction to underwriters in terms of using scores. Often, in an automated underwriting (AU) environment, the lowest-risk loans are approved by an AU system; and a score-whether it be a mortgage or credit score-is a key determinant of that approval. In these settings, it is important that all loans not receiving automated approval be reviewed by underwriters. Furthermore, it is imperative that underwriters clearly be given the latitude to approve a loan even though it has a low mortgage score or the applicant has a low credit score.

Often, AU and scoring systems will flag a loan's strengths and weaknesses, drawing the attention of the underwriter to the portion of the loan that needs the greatest review. This focuses the underwriter's efforts, making him or her more effective and efficient.

At MGIC, we use credit and mortgage scores, but our underwriters also review loan documentation, including credit reports, on loans not instantly approved by our Advance Technology Loan Assessment System (ATLAS). The combination of scoring with ATLAS gives our underwriting function a powerful productivity tool. It improves efficiency and consistency in underwriting execution for less difficult loans while freeing up underwriters to concentrate on loans that require more attention and expertise.

If an underwriter deems that a loan should be rejected, a second underwriter is generally called on to review the file. Either way-automated or manual underwriting-scores are used only to provide direction, not the decision.

We feel this is an appropriately cautious way to proceed into the realm of scoring and automated underwriting while still reaping the benefits of the tools helping to put the industry's growing knowledge of mortgage risk to work. It also is appropriate given that our industry's understanding of what leads to foreclosure is still evolving. The quest for precision, after all, is far from over. M B

Author Affiliation: Larry Pierzchalski is executive vice president of risk management at Mortgage Guaranty Insurance Corporation, Milwaukee, Wisconsin.

Special Features: Charts; Graphs
Company Names:
Mortgage Guaranty Insurance Corp **Duns:** 00-794-7096

Fair Isaac & Co Inc **Duns:** 07-466-4715 **Ticker:** FICI

Classification: 3200 (CN=Credit management); 8120 (CN=Retail banking); 9130 (CN=Experimental/Theoretical); 9190 (CN=United States)
Descriptors: Credit scoring; Models; Comparative **analysis**; Trends; Mortgage banks
Geographic Names: US

36/9/13 (Item 13 from file: 268)
00295985
Fixed-premium deposit insurance and international credit crunches
Spiegel, Mark M
Federal Reserve Bank of San Francisco Economic Review , v 2 , p 3-15 , 1996 **Document Type:** Journal Article **ISSN:** 0363-0021 **Journal Code:** FSE **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 03202

Abstract:

A model of a monopolistically competitive model of foreign lending is presented. In this model, both explicit and implicit fixed-premium deposit insurance increase the degree to which bank participation in relending to problem debtors falls below its globally optimal level. This provides a channel for fixed-premium deposit insurance to inhibit extension in bad states, resulting in an increase in the expected default percentage and an increase in the expected burden on the deposit insurance institution.

Text:

Headnote:

This article introduces a monopolistically competitive model of foreign lending in which both explicit and implicit fixedpremium deposit insurance

increase the degree to which bank participation in relending to problem debtors falls below its globally optimal level. This provides a channel for fixed-premium deposit insurance to inhibit credit extension in bad states, resulting in an increase in the expected default percentage and an increase in the expected burden on the deposit insurance institution.

While the perverse incentives faced by banks due to fixedpremium deposit insurance have been well-documented,1 the literature has largely ignored the potential of deposit insurance to distort the organization of the banking industry. This paper introduces a simple model to fill this gap, in which fixed-premium deposit insurance plays a role in determining the structure of bank lending. To the extent that banking organization affects the ability of banks to act in concert, the paper introduces a new channel through which deposit insurance may have an adverse impact on lending outcomes.

The impact of deposit insurance on bank behavior has long been a source of concern to policymakers and researchers. A large literature exists which argues that fixedpremium deposit insurance increases the riskiness of bank lending portfolios (Kareken and Wallace 1978, Kareken 1986, Penati and Protopapadakis 1988, Jaffee 1989, Kane 1989, Duan, et al., 1992). In addition, Penati and Protopapadakis argue that "implicit deposit insurance," where regulators merge rather than close failing banks out of concern for the stability of the **financial** system, provides an additional subsidy. For example, from 1978 through 1984, only 20 percent of failed U.S. banks were closed. Moreover, these were largely small banks, representing only 0.2 percent of total deposits.

This paper demonstrates that the introduction of fixedpremium deposit insurance, both explicit and implicit, can magnify the degree to which credit extension is sub-optimal by increasing the number of banks participating in the lending package. The **analysis** is conducted through a monopolistically competitive two-period model of foreign lending, introduced in Section II.

The interesting decision in the two-period model comes at the end of the first period. Banks are confronted with the ability to increase the performance of their outstanding loans by rolling over debt at terms that would not appear to be profitable to unexposed creditors. However, there are positive spillovers associated with new lending, which implies that the disparity between the magnitude of new lending and that which is globally optimal will be increasing in the number of exposed banks. The incentive to avoid this public good problem limits the number of participating banks in equilibrium. However, fixed-premium deposit insurance mitigates this incentive, increasing the number of participating banks and exacerbating the public good problem. This reduces the expected percentage of debt service, increases the probability of bank failure, and increases the expected burden on the deposit insurance institution. Simulation results below indicate that fixed-premium deposit insurance can have a relatively large impact on default probabilities.

While our qualitative results apply equally well to any situation where externalities among creditors may exist,2 foreign lending provides a particularly clean example of lending externalities across banks. In domestic lending situations, creditors can partially deal with anticipated future renegotiation difficulties through debt covenants and other contract instruments3 which are not binding in an international lending context. In addition, rescheduling negotiations in international lending take place under the auspices of the Paris Club, which applies the constraint of equal sharing rules, such that all loans have equal seniority. Finally, episodes of perceived sub-optimal credit extension are well documented in international lending, such as the failure of the Baker Plan to deal with the Latin American debt crisis (Cline 1989).

There is considerable evidence that the deposit insurance subsidy affected bank incentives in foreign lending. Event studies of the impact of the debt crisis in August 1982 on bank equity showed a consistently lower impact on excess returns than would be suggested by the magnitude of the

news. For example, while uninsured bond spreads over LIBOR soared from 2 percent in August 1982 to over 7 percent in November (Edwards 1986), there was less than a 2 percent decline in the average annual excess returns of banks exposed to Mexico (Schoder and Vankudre 1986, Bruner and Simms 1987, and Spiegel 1992).4 Similarly, James (1990) finds that changes in the **value** of bank stock equity are smaller than exposure-weighted movements in the secondary market prices of sovereign debt would imply.

This paper is organized as follows: Section I reviews the performance of commercial banks under the Baker Plan. Section II then introduces our theoretical model. The impact of deposit insurance, both explicit and implicit, is examined in Section III. Section IV contains simulations concerning the empirical predictions of the effects discussed in the theory. Section V concludes.

I. THE BAKER PLAN: 1986-1988

The "Baker Plan," named for former Treasury Secretary James Baker, provides one of the best recent examples of collective action problems among international creditors. Subsequent to Mexico's suspension of payment on its external debt, a number of countries experienced difficulties in obtaining **financing**. The general belief concerning the difficulties faced by these countries was that their problems were ones of "illiquidity" rather than "insolvency." In other words, if **financing** could be obtained to get countries through a relatively difficult period, they would then be able to service all of their debts. Subsequent to this period "... countries could grow their way out of debt and could expand their exports enough to reduce their relative debt burdens to levels compatible with a return to normal credit market access" (Cline 1989 p. 177).

The Baker Plan called for commercial banks to extend approximately $7 billion annually, or 2.5 percent of total exposure, to fifteen highly indebted developing countries.5 Cline (1989) claims that banks achieved capital flows of approximately $13 billion over the Baker Plan period, or about two-thirds of their $20 billion target.6 It was well understood at the time that the anticipated disbursements from commercial banks under the plan were by no means certain. Brainard (1987) suggested that banks needed to understand how the involved government intended to manage the Baker Plan or "... increased official lending will merely substitute for reduced bank credits."

In retrospect, the magnitude of actual flows during the Baker Plan period seems even lower than Cline's estimate. Husain (1989) points out that while the IMF estimates that commercial banks committed $16.4 billion in new money and actually disbursed $15 billion-figures the commercial banks themselves used to support their claims of having come close to the Baker Plan targets-debtor country data show that net new long-term **financing** to the highly indebted countries amounted to only $4 billion (see Table

1). Moreover, if private nonguaranteed debt is taken into account, there were net repayments to commercial banks amounting to $2.4 billion. In no case did commercial banks provide more net **financing** than they received in interest payments. Husain also adds that "U.S. banks have been most active in reducing their developing country exposure. Between mid-1987 and the end of the third quarter of 1988, these banks reduced their claims on all developing countries by more than $20 billion. More than half of this represented a reduction in claims on highly indebted countries" (p. 14).

In summary, the Baker Plan is an example of collective action problems across commercial banks. It was generally perceived that increasing exposure in the aggregate was desirable from the point of view of the exposed banks, but individually each bank had the incentive to "free-ride" on the efforts of the other creditors by not fulfilling its disbursement commitment. The result was that the level of new money extended to the highly indebted nations was sub-optimal from the aggregate creditor perspective. Because commercial banks faced incentive problems in taking

collective action, a resolution of the debt crisis required turning to voluntary methods, such as the "market-based menu approach" associated with the Brady Plan (see Diwan and Spiegel 1994).

II. A MONOPOLISTICALLY COMPETITIVE MODEL OF BANK LENDING

Setup

In this section, we derive a formal model which exhibits collective action problems similar to those encountered under the Baker Plan. There are three types of players in the model, a debtor nation government, a group of identical monopolistically competitive banks, and atomistic depositors. The extensive form of the model has five stages: In the first stage, the magnitude of first-period lending by individual banks, l sub1 , and the number of banks participating in the initial lending package, n, are determined. Total first period lending, L sub1 , then equals nl sub1 . Loans are assumed to be short term, coming due at the end of the first period.7 For simplicity, all first-period lending is assumed to be consumed. First-period output of the debtor nation, ql, is determined in the second stage. In the third stage, banks choose an amount of new lending l sub2 , so that total new lending L sub2 = nl sub2 . In the fourth stage, the debtor nation government chooses its percentage of debt service on outstanding first-period loans, pi sub1 , where pi sub1 element [0,1]. Finally, second period debtor nation output, q sub2 , is determined in the final stage, which simultaneously determines the percentage of debt service on outstanding second period lending, pi sub2 , where pi sub2 element [0,1].

q sub1 and q sub2 are assumed to be exogenous independently distributed **random** variables distributed uniformly on the **interval** [0,1]. sup8 Total lending in each period is equal to the number of banks participating in the lending package, n, times individual bank lending,l sub 1 . It. As we show below, n will be constant across periods. Let r sub1 represent one plus the contractual nominal rate of interest on the loan in period iota (iota = 1, 2). For simplicity, r sub t is taken as exogenous.9 Consequently, the outstanding obligation on period t loans at the end of period t will be equal to r sub t nl sub l .

V. CONCLUSION

In this paper, we examined the implications of fixed-premium deposit insurance in a foreign lending model where rescheduling exhibits positive spillovers across creditors. Our results show that deposit insurance raises the number of banks participating in the lending package through three channels: First, deposit insurance acts as a subsidy on lending; second, deposit insurance weakens the degree to which the market induces banks to organize in a manner that will minimize the public good problem associated with relending to a problem debtor; finally, implicit deposit insurance removes much of the remaining liability side of the bank balance sheet from a private regulating role. Moreover, if the deposit insurance institution's appraisal of the degree of systemic risk in a lending package is endogenous, banks will be rewarded for organizing themselves in a manner that enhances the probability of a bail-out.

Both private creditors and government officials of lending and borrowing countries have argued that the level of loan provision to the highly indebted countries during the debt crisis was sub-optimal from the point of view of the industry as a whole. Previous discussions explain underlending through "herd behavior" followed by banks (Herring and Guttentag 1985). This paper shows that sub-optimally large levels of banking "diffusion," rationally introduced to avoid firm risk and take advantage of fixed-premium deposit insurance, may exacerbate the degree to which credit extensions are sub-optimal, providing an alternative explanation to herd behavior.

(Table Omitted)

Captioned as: TABLE 1

(Graph Omitted)

Captioned as: FIGURE

(Graph Omitted)
Captioned as: FIGURE 2
(Table Omitted)
Captioned as: TABLE 2
Footnote:
1. See Santomero (1984) for an extensive survey of this literature.
Footnote:
2. See Bernanke (1991) for a discussion of domestic "credit crunches."
3. See, for example, Berlin and Mester (1992).
4. Beebe (1985) and Cornell and Shapiro (1986) show larger declines over the long run, but their measures do not approach the magnitudes observed in bond-spread movements.
Footnote:
5. The original countries were Algeria, China, Egypt, Greece, Hungary, India, Indonesia, Malaysia, Pakistan, Poland, Portugal, Thailand, and Turkey; Costa Rica and Jamaica were added later.
6. Ironically, public sector capital flows to the Baker nations actually did worse over the plan, falling by $4 billion annually. This occurred because of large decreases in IMF and bilateral lending.
Footnote:
7. Creditors might respond to future anticipated renegotiation problems by lengthening the maturity of their debt contract (Sharpe 1991). In practice, however, creditors responded to the increase in the perceived riskiness of the highly indebted countries by shortening the maturity of their loans. The lack of long-term lending may stem from equal-sharing, since some of the benefits of extending a long-term loan would spill over to short-term creditors.
8. We assume that q sub1 and q sub2 are independent for simplicity. This does not drive the results below, but it does increase the parameter space in which relending to a problem debtor is an optimal response. If the two were positively correlated, it would be more "likely" that relending would be throwing good money after bad. Nevertheless, since the point of the exercise is to introduce an example where collective action problems may arise, this assumption is relatively innocuous.
Footnote:
The assumption of exogeneity of these output variables is also made for simplicity and drives none of the results below. However, allowing for "debt overhang" effects, where indebtedness may affect output levels, may also affect the desirability of relending in a more general model.
9. It is well known that allowing banks to choose both It and rt would result in a multiplicity of equilibria. Holding rt constant is valid if the debtor is credit constrained, which we assume for a problem international debtor.
Footnote:
10. The default penalty is needed to generate positive lending in equilibrium, but its specification does not drive our results. Default penalties in sovereign lending have been motivated by loss of future access to capital markets (Eaton and Gersovitz 1981), seizure of assets (Bulow and Rogoff 1989), or loss of reputation (Grossman and Van Huyk 1988). Lindert and Morton (1989) show that the ex-post rate of return on sovereign lending has **historically** been competitive, implying that the perception of a penalty for default exists.
Footnote:
11. Generalizing the model by allowing depositors to increase tau by "brokering" deposits across a number of banks would actually strengthen the results below by providing an additional incentive for an increase in the number of banks.
Footnote:
12. Equation (7) is simplified by noting that on the margin d(pi) sub t ,/dl sub 2 = d/aL (t = 1, 2).
Footnote:
13. Note that the impact of changes in yi on the probability of

bankruptcy does not affect yr, because they already reflect a liability of the deposit insurance institution.

Footnote:

14. Penati and Protopapadakis (1988) suggested alternatively that the probability that loans are implicitly insured may be increasing in the number of banks involved in the lending package, so that ps = ps(n). The qualitative results under this alternative criterion would be identical.

15. Since 0<pp,<l, this linear specification must be considered as a local approximation to a non-linear function.

Footnote:

16. This share corresponds to that which existed on average from 1980 to 1985 according to Penati and Protopapadakis (1988).

17. These parameters have been chosen to insure that 0<iota<1.

18. The surprising result that the number of banks in the system actually declines with increases inphistems from the zero-profit condition. Since increases inphimake lending less profitable, and individual bank lending remains constant, exit must occur for profits to return to zero.

Reference:

REFERENCES

Reference:

Beebe, J. 1985. "Bank Stock Performance Since the 1970s." Federal Reserve Bank of San Francisco Economic Review 5, pp. 5-18. Berlin, M., and L. J. Mester. 1992. "Debt Covenants and Renegotiation."

Journal of **Financial** Intermediation 2, pp. 95-133.

Bernanke, B. 1991. "The Credit Crunch." Brookings Papers on Economic Activity, pp. 204-239.

Brainard, L. J. 1987. "Managing the International Debt Crisis: The Future of the Baker Plan." Contemporary Policy Issues 5(3), pp. 66-75.

Bruner, R. F., and J. M. Simms, Jr. 1987. "The International Debt Crisis and Bank Security Returns in 1982." Journal of Money; Credit and Banking 19, pp. 47-55.

Bulow, J., and K. Rogoff. 1989. "A Constant Recontracting Model of

Sovereign Debt." Journal of Political Economy 96, pp. 155-178. Cline, W R. 1989. "The Baker Plan and the Brady Reformulation: An Evaluation" in I. Husain and I. Diwan, eds., Dealing with the Debt Crisis. Washington D.C.: World Bank, pp. 17-193. Cornell, B., and A. C. Shapiro. 1986. "The Reaction of Bank Stock Prices to the International Debt Crisis." Journal of Banking and **Finance** lO, pp. 55-73.

Reference:

Diwan, I., and M. Spiegel. 1994. "Are Buybacks Back?: Menu-Driven Debt Reduction Schemes with Heterogeneous Creditors." Journal of Monetary Economics 34, pp. 279-293.

Duan, J. C., A. F. Moreau, and C.W Sealey. 1992. "Fixed-Rate Deposit Insurance and Risk-Shifting Behavior at Commercial Banks." Journal of Banking and **Finance** 16, pp. 715-742. Eaton,J., and M. Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical **Analysis**." Review of Economic Studies 83, pp.289-309.

Edwards, S. 1986. "The Pricing of Bonds and Bank Loans in International Markets." European Economic Review 30, pp. 565-589. Grossman, H., and J. B. Van Huyk. 1988. "Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation." American Economic Review 78, pp. 1088-1097. Herring, R. J., and J. M. Guttentag. 1985. "Commercial Bank Lending to Developing Countries: From Overlending to Underlending to Structural Reform." Wharton School International Banking Center. Husain, 1. 1989. "Recent Experience with the Debt Strategy." **Finance** and Development 26(3) September, pp. 12-15.

Reference:

Jaffee, D. M. 1989. "Symposium on Federal Deposit Insurance for S&L Institutions." Journal of Economic Perspectives 3, pp. 3-10. James, C. 1990. "Heterogeneous Creditors and the Market **Value** of Bank LDC Loan Portfolios." Journal of Monetary Economics 25, pp. 325-346.

Kane, E. J. 1989. "The High Cost of Incompletely Funding the FSLIC

Shortage of Explicit Capital." Journal of Economic Perspectives 3, pp. 31-47.

Kareken, J. H. 1986. "Federal Bank Regulatory Policy: A Description and Some Observations." Journal of Business 59, pp. 3-48. , and N. Wallace. 1978. "Deposit Insurance and Bank Regulation: A Partial Equilibrium Exposition." Journal of Business 51, pp. 413-438.

Reference:

Lindert P., and P Morton. 1989. "How Sovereign Debt Has Worked" in J. Sachs, ed., Developing Country Debt and the World Economy. Chicago: University of Chicago Press, pp. 225-236. Penati, A., and A. Protopapadakis. 1988. "The Effect of Implicit Deposit Insurance on Banks' **Portfolio** Choices with an Application to International 'Overexposure'." Journal of Monetary Economics 21, pp.107-126.

Santomero, A. M. 1984. "Modeling the Banking Firm: A Survey." Jour

nal of Money, Credit and Banking 16, pp. 576-602. Schoder, S., and P Vankudre. 1986. "The Market for Bank Stocks and Banks' Disclosure of Cross-Border Exposure: The 1982 Mexican Debt Crisis." Studies in Banking and **Finance** 3, pp. 179-202. Sharpe, S. A. 1991. "Credit Rationing, Concessionary Lending, and Debt Maturity." Journal of Banking and **Finance** 15, pp. 581-604. Spiegel, M. M. 1992. "Concerted Lending: Did Large Banks Bear the Burden?" Journal of Money, Credit and Banking 24, pp. 465-482.

Author Affiliation:

Mark M. Spiegel

Senior Economist, Federal Reserve Bank of San Francisco. Helpful comments were received from Tim Cogley, Ken Kasa, Simon Kwan, Mark Levonian, and Anthony Venables.

Special Features: Charts; Graphs; Equations; References
Classification: 8110 (CN=Commercial banking); 1300 (CN=International trade & foreign investment); 9130 (CN=Experimental/Theoretical); 9190 (CN=United States)
Descriptors: International lending; Deposit insurance ; Studies
Geographic Names: US

36/9/14 (Item 14 from file: 268)
00276415
Performance contracts for central bankers
Waller, Christopher J
Federal Reserve Bank of St. Louis Review , v 77 , n 5 , p 3-14 , Sep/Oct 1995
CODEN: FRBRDV **Document Type:** Journal Article **ISSN:** 0014-9187 **Journal Code:** FSL **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 07308

Abstract:

Discretionary monetary policy may lead to an inflationary bias, which arises when the target level of output is higher than the trend **value**. There have been 2 distinct directions of research on ways of eliminating this inflationary bias: the reputation-building approach and the institutional design approach. Recently, a new idea has

surfaced in the institutional-design literature for dealing with the inflation bias. The idea is to offer the central banker a performance contract whereby the central banker's salary or the bank's budget is tied directly to the performance of important macroeconomic variables such as GDP and the inflation rate. By giving the central banker the proper **financial** incentives, researchers have shown that the central bank can be induced to generate low inflation without forsaking its stabilization responsibilities.

Text:

Since the end of World War II, economies around the world have been plagued by **historically** high and persistent inflation. This raises a question: If inflation is socially undesirable, why do policymakers produce it? One explanation is that discretionary monetary policy may lead to an inflationary bias. This explanation is based on the "time-inconsistency" problem, first outlined by Kydland and Prescott (1977) and illuminated by Barro and Gordon (1983a). The typical version of this explanation assumes that society wants the monetary authority to follow a low inflation policy, which it promises to do. Once private agents commit themselves to nominal wage contracts based on a low expected inflation rate, however, the monetary authority is assumed to have an incentive to create "surprise" inflation and inflate away the real **value** of the contracted nominal wage. As a result, firms hire more labor and produce more output. But, because private agents are aware of this incentive, they do not believe that the central bank will carry through with its promise to maintain inflation at a low level. Hence workers set their nominal wages high enough so that the extra inflation created by the central bank leaves real wages at their desired levels. Consequently, no additional output or employment is created but society suffers from an inflation bias.

For the past decade, researchers have investigated an array of methods with which to reduce this inflation bias. Although most methods promise to lower the inflation bias, they usually do so at the cost of creating greater output variability. However, a recent proposal by Walsh (1995a) and Persson and Tabellini (1993)--the adoption of performance contracts for central bankers--has created a stir among economists working in this area. The purpose of this article is to survey the work on performance contracts and compare it to earlier proposals for mitigating the inflation bias. The remainder of the paper proceeds as follows: The second section contains a model describing the basic time-inconsistency problem and reviews previous suggestions for eliminating the inflationary bias. Following that is a discussion of the nature of performance contracts and how they work. The fourth section probes the principal-agent nature of central banking and its relationship to central bank independence. In the final section, I offer concluding comments.

THE TIME-INCONSISTENCY PROBLEM

A Model of Discretionary Monetary Policy

A general description of how monetary policy is determined would go something like this: Society (the principal) delegates the power to create money to the central bank (the agent). Society instructs the central bank to use its money creation powers to "do good." What is meant by doing good is often not well-defined: nevertheless. it can be interpreted to mean that the central bank should produce a policy that improves the well-being of society. The central bank then enacts policy according to some objective function. Presumably, its objective is to maximize social welfare. but it could also be to maximize something other than society's welfare. Finally, after policy is enacted, the monetary authority may be asked to account for its actions.

To illustrate the nature of the time-inconsistency problem, consider the following version of the Barro and Gordon model:

(Equation 1 omitted)

(Equation 2 omitted)

(Equation 3 omitted)

where y is real output, y sup n is the trend level of output, pi is the inflation rate, pi sup e is the expected inflation rate and u is a mean zero, serially uncorrelated real outputshock.

Equation 1 describes how output is influenced by inflation and inflation expectations. Workers are assumed to sign nominal wage contracts prior to the setting of monetary policy and the contracted wag is based on the expected rate of inflation. An inflation surprise reduces the real **value** of the contracted nominal wage, thereby inducing firms to hire more labor and produce more output.

Equation 2 is society's utility function and shows that society suffers from output and inflation fluctuations about their targeted levels. Society's target output level is y sup n + k sup S where y sup n is the natural or trend level of output and k sup S is a positive constant. The parameter k sup S is assumed to reflect society's belief that distortions in the economy make trend output undesirably low. Society's preferred inflation rate is assumed to be zero. The parameter b measures the relative weight society places on losses arising from inflation. The weight on losses arising from output has been set equal to 1 for notational ease.

Equation 3 is assumed to be the central banker's objective function. The parameter w is the salary or budget the central banker receives for doing the job. This term is irrelevant in the standard Barro and Gordon model and is usually ignored. But this term plays a key role in the performance-contract literature, so I will include it now for comparison later. Equation 3 looks very much like society's utility function except that the central bank is allowed to have a potentially different output target, y sup n + k sup M , than society's. If k sup M = k sup M , then the central bank's objective is identical to society's. If k sup M -= k sup M , then the central bank uses policy to pursue an agenda that is different than that of society as a whole. The reason the central bank has a different agenda is important and is a crucial part of the performance-contract debate, as discussed later in this article. Finally, for ease of **analysis**, the monetary authority is assumed to control the inflation rate directly and thus chooses pi to maximize equation 3 given equation 1.

Consider the case in which the central bank has only society's interests at heart, that is, k sup M = k sup M = k. Since society wants inflation to be zero (on average), suppose the central bank can pre-commit to a policy whereby it will not create systematic inflation. This implies that expected inflation is zero. Substituting equation 1 into 3 and maximizing subject to the constraint k sup M = k sup S = 0 yields what is called the socially optimal or "pre-commitment" solution for inflation and output:

(Equation 4 omitted)

(Equation 5 omitted)

From equation 4, the central banker partially offsets the output shock by allowing inflation to vary more. Expected inflation is zero, and expected output is y sup n . In this world, pre-commitment refers to the idea that the central bank can commit itself to making the inflation rate zero on average, but will vary the period-by-period inflation rate to stabilize output in a way that maximizes social welfare. The central bank makes no attempt to expand output above the trend level even though it has a desire to do so. In short, even though k > 0, pre-commitment means the central bank is able to credibly promise to act as if k = O.

Now suppose that the central bank cannot commit itself to acting as if k = O. Now the central banker chooses pi, taking pi sup e as given, to maximize its objective function. Maximizing equation 3 yields the following expression:

(Equation 6 omitted)

Rational expectations implies that pi sup e must be set consistent with equation 6. This implies that

(Equation 7 omitted)

which yields the following solutions for the discretionary equilibrium:

(Equation 8 omitted)

(Equation 9 omitted)

The only difference between these expressions and those from the pre-commitment solution is that there is now an inflationary bias, given by k sup M /b > 0; output is the same.

Why does the inflation bias arise? Because the target level of output is higher than the trend **value**. Once wage contracts are signed, the central bank can increase output above trend by creating an inflation surprise. The central bank does this not out of self-interest but because society wants it to. Even though society as a whole desires this, however, individual agents have no incentive to allow their wages to be inflated away. Consequently, they set expectations and nominal wage demands accordingly. In equilibrium, the economy suffers from excessive inflation with no additional gains in output. It can be shown that the loss from the discretionary equilibrium is higher than it would be in the pre-commitment case. Thus, even though the central banker does what society wants him to do, the use of discretionary policy makes society worse off in equilibrium.

There are three points to note about equations 8 and 9. First, the inflation bias is a constant--it is not a **random** variable nor does it vary over time. Second, the bias does not depend on the output shock. Third, the stabilization response to the output shock u is the same in both the socially optimal solution and the discretionary solution. These features all come into play when discussing the optimal design of performance contracts.

RESOLVING THE TIME-INCONSISTENCY PROBLEM

Since the publication of the Barro and Gordon (1983a) paper, research has focused on ways of eliminating this inflationary bias. There have been two distinct directions of research: the reputation-building approach and the institutional-design approach.

Reputation Building

The reputation-building approach focuses on the use of "punishment" strategies by private agents to deter the central bank from generating the inflation bias. In these models, workers believe the central bank will follow a low inflation policy as long as it has not tried to surprise workers in the past. Otherwise, they "punish" the central bank by expecting a high inflation rate, which the central bank validates to avoid creating a recession. By using this type of mechanism, the private sector is able to persuade the central banker to develop a reputation for enacting the announced policy. Barro and Gordon (1983b) showed that reputation building would generate a lower inflation bias but would not eliminate it.

Barro and Gordon's early model of reputation was done under the assumption of perfect information. Subsequent research examined how robust the reputation building approach was to information imperfections. Canzoneri (1985) showed that the economy would suffer inflation "cycles" due to occasional breakdowns in credibility if private agents were unable to separate exogenous inflation shocks from systematic policy actions. Backus and Driffill (1985), Barro (1986) and Rogoff (1987) showed that if private agents are unsure of the central banker's type--inflation hawk or dove--then a recession will frequently occur early in a central banker's term. This is because private agents' expectations of inflation are an average of the hawk's and the dove's equilibrium inflation rates. If the central banker is a hawk, inflation is set lower than expected and a recession occurs. If the central banker is a dove, he may act like a hawk and create a recession to build a reputation as a hawk. The reason is that if the dove inflates immediately, he reveals himself as a dove and inflation expectations will be higher for the remainder of his term in

office. By acting like a hawk, he manages to keep inflation expectations low. The dove, however, eventually chooses to create an inflation surprise and expand output for a short period of time. Thus, while inflation is lower on average, output and inflation are more variable.

Although reputation models are able to generate lower equilibrium inflation rates, albeit at some cost of greater output variability, they have several unappealing aspects. First, there are an infinite number of punishment strategies that could be adopted, and it is not obvious which is the correct one to use. For example, how long should the punishment last?

Second, the multiplicity of strategies suggests that private agents would have to coordinate their actions to send a clear signal to the central bank as to how they would behave in the event that they are surprised. But how is such coordination to be achieved? Large, national trade unions may be sufficient for coordinating actions in some countries, but this is not a feasible solution in the relatively atomistic labor markets that characterize the U.S. economy.

Third, the reputation approach tends to focus on the personality and reputation of individual central bankers. Because individuals do not serve as the central banker for long periods in the real world, this approach suggests that there will be considerable uncertainty and variability of policy as central bankers turn over. Thus, we should focus on ways of developing the institutional reputation of the central bank instead of the reputation of individual central bankers.

Finally, the reliance on the private sector to enforce the appropriate path of monetary policy is a bit unpleasant from a public policy perspective. The reputation--building approach does not try to change the central bank's objective function directly; rather, it alters the central bank's behavior by making the policy choice dynamic, that is, by making today's policy actions have future consequences. But if the institutional structure of the central bank provides it with the wrong policy incentives, then it would seem prudent to change the institution rather than rely on private agents to solve the problem.

To illustrate this point, consider the response to airline hijackings. One way of dealing with hijackers is to arm the passengers and let them enforce peace on the airplane. This is akin to what the reputation approach does for the inflation bias. A better idea is to change the environment for boarding a plane so that the likelihood of a hijacking is reduced--hence, the use of metal detectors.

As a result of these problems with the reputation--based approach, researchers began to investigate institutional reforms for the central bank that would mitigate the inflationary bias.

Institutional Design

The institutional--design approach focuses on using legislative means to restrain the central bank from engaging in high-inflation policies. The intent is to manipulate the central bank's objective function directly through legislative action. Some work in this area has focused on legislation that restricts the day-today operating procedures of the central bank; other research shows how the appointment process for central bankers can be used to elicit better inflation performance. Advocates of the latter line of research recommend making the central bank independent from elected leaders as a means of reducing the inflationary bias.

Targeting Regimes

Legislative restrictions on the central bank often take the form of imposing monetary targeting or adopting simple rules (which are actually targeting regimes with a horizon of one period). The adoption of Friedmanesque k-percent rules has been studied by Alesina (1988) and Lohmann (1992). They show that these rules eliminate not only the inflationary bias, but also stabilization of output by the monetary authority. Hence, there is a trade-off between reducing inflation and stabilizing output. Simple rules dominate discretion when output shocks are small and relatively rare.(1)

Multi-period targeting horizons have been examined by Canzoneri (1985)

and Garfinkel and Oh (1993). In these models, the central bank must follow policies so that the targeted inflation rate occurs on average over some time **interval**. In this environment, the central bank creates an inflation bias early in the targeting horizon, but it is smaller than it would have been in the absence of targeting. However, it produces sub-optimally low inflation (or even deflation) at the end of the targeting horizon to hit the targeted inflation or money growth rate. Stabilization is also sacrificed in the name of inflation, since shocks early in the period are not stabilized in an optimal fashion because those actions must be reversed later in the targeting period.

An implicit assumption in these targeting models is that the central banker's worst penalty for missing the target is dismissal (shooting him is not a realistic punishment). Consequently, the central banker's self-interest plays a large but hidden role in these types of models.

Conservative Central Bankers

The appointment and reappointment of a central banker who sets policy according to his own self-interest plays a large role in other institutional schemes for dealing with the inflation bias. Thompson (1981) and Rogoff (1985) proposed appointing a "conservative" central banker who dislikes inflation more than everyone else in society. A conservative central banker generates a lower inflationary bias but does so by not stabilizing the economy in a socially optimal fashion.(2) To illustrate this point, suppose that society appoints a central banker who puts more weight on inflation than it does. The central banker would then have a larger **value** of the parameter b in equation 3 to use in setting policy. From equations 8 and 9, however, we see that a larger **value** of b reduces the inflation bias but makes output more variable.

For the conservative central banker's policies to be credible, society must believe that he cannot be removed ex post by the current government. Thus, the central banker must have some degree of independence to pursue policies that are not desired by the current administration (and, implicitly, the electorate). Subsequent research by Flood and Izard (1989) and Lohmann (1992) showed that complete independence was not socially optimal--for certain bad states of the world, society benefits from firing the conservative central banker and stabilizing output.

A NEW INSTITUTIONAL DESIGN: PERFORMANCE CONTRACTS

A consistent theme of both the reputation-building and institutional--design models is that the inflation bias can be reduced or eliminated, but usually at the cost of having the central bank reduce its emphasis on stabilizing output. Thus, there appears to be a trade-off between reducing average inflation and stabilizing the real economy.(3) Debate has centered on the relative benefits and costs of this trade-off in determining the goals of monetary policy, and the types of legislative restraints to place on the central bank.

Recently, however, a new idea has surfaced in the institutional-design literature for dealing with the inflation bias. The idea is to offer the central banker a performance contract, whereby the central banker's salary or the bank's budget is tied directly to the performance of important macroeconomic variables such as GDP and the inflation rate. By giving the central banker the proper **financial** incentives, these researchers have shown that the central bank can be induced to generate low inflation without forsaking its stabilization responsibilities.

Performance Contracts

Walsh (1995a) suggested that the monetary policy game be viewed as a principal-agent problem.(4) In a principal--agent model, one individual or group (the principal) delegates control over a policy variable to another individual or group (the agent). Although the principal would like the agent to set policy so that the principal's welfare is maximized, the agent has a different objective and opts for a policy that does not give the principal its most desired outcome. The solution to this problem is for the principal to offer the agent a contract that gives the agent the incentives

to enact the policy desired by the principal.

By viewing monetary policy as a principal-agent model, Walsh redirected attention to the source of the problem--the central banker is confronted with a set of preferences that do not yield the outcome that society prefers most. So rather than worry about appointing conservative central bankers or adopting appropriate reputation strategies, Walsh argued that we should provide the central banker with the incentives to "do the right thing"--even if those incentives do not appear, at first glance, to be consistent with maximizing society's well-being. The problem is determining what those incentives should be.

Following the principal-agent literature, Walsh proposed offering the central bank a performance contract. This contract ties the central banker's personal compensation or the size of the bank's budget to the performance of the economy. Once the contract is signed, society encourages the central banker to pursue his own self-interest and adopt policies that increase his income or the bank's budget. The trick is to structure the contract in such a way that by trying to increase his own resources, the central banker maximizes social welfare in the process.

This approach is a radically different way to deal with policymakers. Under this institutional design, society exploits the pursuit of self-interest by the central banker to achieve the socially desirable outcome. This differs from the traditional view of appointing a benevolent central banker and then instructing him to do good. Under the performance contract approach, society essentially says: "You can do what you want, but you will pay personally for undesirable outcomes." Making the central bank accountable for its actions is a prominent theme of performance contracts.

Designing a Performance Contract

What does a performance contract look like? Consider the following compensation contract for setting the central banker's salary (w in equation 3):

(Equation 10 omitted)

where s denotes the central banker's base salary or the budget of the central bank. This contract specifies that the central banker be paid a base salary s, which will be reduced if any inflation occurs. The degree of salary reduction is determined by the parameter lambda. A key feature of this contract is that it is based solely on the publicly observed inflation rate; it is not based on items that are unverifiable (such as how hard the central banker works).

Once the contract is in place, society tells the central banker to set policy in any manner he sees fit; there is no mention of pursuing the public good. Therefore, given equations 1, 3 and 10, the central banker chooses pi to maximize

(Equation 11 omitted)

This yields the following expression for the inflation rate:

(Equation 12 omitted)

Imposing rational expectations yields the following equilibrium solutions for inflation and output:

(Equation 13 omitted)

(Equation 14 omitted)

Given these expressions for what inflation and output will be when the central banker pursues his own self-interest, society would like to set the weight lambda such that the expressions in 13 and 14 are exactly the same as those given by the precommitment solutions in equations 4 and 5. This result can be accomplished by setting:

(Equation 15 omitted)

By setting lambda = k sup M , the reduction in salary from creating an inflation surprise just offsets any benefits that would accrue from expanding output towards y sup n + k sup M . Hence, on the margin, the loss of income for the central banker is just equal to the utility gain from creating surprise inflation and expanding output, so he chooses not to create surprise inflation and no inflation bias occurs.

Furthermore, output and inflation are stabilized in the socially

optimal fashion. The reason this can be accomplished is that the inflation bias is constant and independent of the output shock u. So a simple linear penalty for inflation is sufficient to deter the central bank from inflating. But the key point is that eliminating the inflation bias through appropriate incentives does not require the central banker's stabilization response to be distorted. Therefore, there is no cost for eliminating the inflation bias. By careful construction of the central banker's compensation, society is able to eliminate the inflation bias and have output optimally stabilized. This is indeed a pleasant result.

The contract could take a variety of different forms and still generate the optimal outcome. Every contract, however, must have the feature that the central bank pays more attention to inflation (or less attention to output) than society does. This simply reflects Rogoff's (1985) notion of a conservative central banker. The only difference is that in Rogoff's framework, society carefully selects a central banker who has the "right" personal attributes to reduce inflation, whereas the contract approach gives any arbitrarily chosen central banker the appropriate incentives to produce low inflation. In general, the principle of Rogoff's idea is still relevant; the issue is how to define "conservative."

Rogoff's definition of a conservative central banker was someone who put more weight on inflation relative to stable output. But we could define a conservative central banker as someone who has a lower inflation rate target or lower output target than the rest of society. In all cases, the central banker cares relatively more about inflation than output.

For example, consider the following performance contract:

(Equation 16 omitted)

In this example, society simply offers the central banker a contract that penalizes him if output is above the natural rate, plus adds a fixed amount to the base salary according to the magnitude of k sup M . Substituting 16 into 3 and rearranging yields

(Equation 17 omitted)

The contract in 16 leads to an objective function for the central banker that is equivalent to appointing a central banker with a lower output target than the rest of society, since the parameter k sup M disappears. With this contract the central banker will use discretion to produce the socially optimal outcome.

Alternatively, Svensson (1995) proposes a contract of the form:

where pi sup c is an arbitrary constant to be determined by society. Substituting 18 into 3 and rearranging yields

(Equation 18 omitted)

If k sup M = k sup S , this contract looks very much like society's utility function except that the central banker's target inflation rate is now different from zero. Thus, the contract in 18 is observationally equivalent to appointing a central banker with a different inflation target than the rest of society's. A central banker with this contract will set policy such that, in equilibrium, inflation and output are given by

(Equation 2 omitted)

(Equation 21 omitted)

Notice that in setting pi sup c = -k sup M /b, we obtain the socially optimal solution. Thus, by having the central banker target a desired inflation rate of minus the inflation bias, society obtains its most preferred outcome. Because the central banker's targeted inflation rate is less than society's preferred rate, the central banker appears more conservative than the rest of society; in contrast to Rogoff's model, however, this type of conservative central banker does not cause stabilization to be sub-optimal.

The key point of this discussion is that offering the central banker a performance contract may be equivalent to appointing an appropriately defined conservative central banker. Once we realize this, there is no reason to believe that these central bankers will understabilize the economy.

IS TIME INCONSISTENCY THE SAME AS A PRINCIPAL-AGENT PROBLEM?

In the performance contract approach above, it was shown that appropriately chosen contracts can induce the central banker to produce the socially optimal outcome. This result was demonstrated without any reliance on the assumption that the central banker's output target was equal to society's. Walsh conducts his **analysis** under the assumption that society and the central banker have the same objective functions, that is, k sup M = k sup S . This assumption is common in the time-inconsistency literature, but is not consistent with the principal-agent model. Usually in a principal-agent problem, the agent has a different objective than the principal. A more classical depiction of the principal-agent problem would look like the following utility functions:

(Equation 22 omitted)

(Equation 22 omitted)

With this formulation, society has preferences that are consistent with the socially optimal solution given in 4 and 5. The central bank, on the other hand, wants output to be higher than its trend **value** (for some unspecified reason). Thus, the central banker uses his discretionary powers to create an inflation surprise, thereby expanding output. Rational agents foresee this and adjust wages so that they are not fooled. The outcome is an inflation bias with no additional output gains.

Although the story is the same as the time-inconsistency model described above, there is one fundamental difference: Society does not want the central bank to try to expand output above trend. The central bank does so in pursuit of its own self-interest. This situation is what performance contracts were designed for: enticing a "misbehaving" agent to produce the principal's desired policy.

But if the performance contract generates the socially optimal outcome regardless of whether society and the central banker have the same output targets, why is it important to classify the problem as a time-inconsistency problem rather than a principal-agent problem? The reason is that if the policy game is described as the principal-agent problem as in equations 22 and 23 above, the credibility of contract enforcement is not an issue. The principal very clearly wants the socially optimal policy to be implemented and has every incentive to hold the central banker to the contract and not renegotiate it. But in the case in which the central banker is trying to give society what it wants, society is inconsistent--it wants higher output, which can only be achieved by being "fooled;" yet, society does not want to be fooled. If the central banker is maximizing social welfare, then society should renege on the performance contract once private agents set their wages--it should let itself be fooled. Since it is optimal ex post to renege on the performance contract, then private agents will never believe it changes the central banker's incentives, and we are right back where we started.

The credibility of contract enforcement raises an important point: Time-inconsistency and principal-agent relationships are not the same thing, even though performance contracts appear to solve both types of problems. Thus, one needs to be careful in using solution concepts interchangeably.

Enforceability of the performance contract corresponds to McCallum's (1995) second fallacy of central bank independence. McCallum argues that a performance contract "does not actually overcome the motivation for dynamic inconsistency; it merely relocates it" (p. 210). As long as the central banker is presumed to be maximizing social welfare, this argument is correct. But if the inflation bias is actually the result of a "true" principal-agent problem rather than a time-inconsistency problem, society can pre-commit itself to enforcing the contract.

Actually, McCallum's criticism of performance contracts is too strong. While it is correct to say that a perfect commitment technology or institutional design does not exist (for example, even the U.S. Constitution is not a perfect commitment to liberty because we can change it anytime we want), it is possible to make the costs of reneging on promises more costly and thus make monetary policy more credible. The basic

idea of performance contracts, and the premise behind the entire institutional-design literature, is to increase the cost of reneging on a cooperative arrangement. Some institutions have low reneging costs (a policy target), while others have very high reneging costs (abolishing the Fed). By relocating the source of dynamic consistency, performance contracts attempt to increase the costs of reneging on low inflation promises.

A TALE OF TWO PRINCIPAL-AGENT PROBLEMS

In equations 22 and 23, the central banker has different objectives than society as a whole in that he wants to increase output above the current trend **value**. This mathematical form corresponds to the traditional principal-agent problem. But why would the central bank have an objective that differs from what society wants?

The answer to this question lies in the policy structure of most democracies. The general public elects a leader who either conducts policy himself or delegates the control of policy to someone else. Monetary policy typically falls in the delegation category. In the United States, for example, voters elect the President and members of Congress who, in turn, delegate the control of monetary policy to the Federal Reserve. Although they delegate control of monetary policy, the President and the Senate jointly determine who shall serve as the head of the Federal Reserve. Thus, there are typically three actors in any monetary policy model: the voters, the elected leaders and the central banker In the time-inconsistency model, all of these actors are assumed to have the same objective. From a principal-agent perspective, however, the presumption is that they have differing objectives.

The "Rogue" Central Banker

Consider the following principal-agent problem. The voters and elected leaders have the same policy objective, given by equation 22, while the central banker has the objective function given in 23. In this case, the central banker is a "rogue" policymaker who sets policy to maximize his self-interest rather than society's or the elected leaders' and who, by doing so, creates an inflation bias.

Why would the central bank behave this way? Central bankers may want to maximize their amenities such as the number of staff members, the luxuriance of buildings and the size of travel budgets, all of which are funded by excessive seignior-age creation.(5) Or if the central bank is unduly influenced by a special interest group, say the banking/ **financial** sector, it may pursue policies that benefit these sectors rather than society. Regardless of the source of the problem, performance contracts are a desirable way of dealing with it. Society and the elected leaders use a performance contract to rein in the central banker and make him accountable to the electorate (why elected leaders do not simply take control of monetary policy then is somewhat puzzling).

According to this scenario, central bank independence is an undesirable institutional-structure. The performance contract approach can work only if the elected leaders have control over the central bank through the setting of budgets and salaries, and the ability to dismiss the central banker over policy actions. For example, Walsh (forthcoming) shows that if adjusting the bank's budget and salaries is infeasible, then threatening to dismiss the central banker if certain poor policy outcomes arise can replicate the equilibria supported by performance contracts. Walsh refers to these optimally designed threats as "dismissal contracts," since the central banker knows exactly which conditions will lead to his dismissal and agrees to such an arrangement.

The implications for central bank independence in this setting are very different from what is generally thought to be. Central bank independence is generally believed to be a crucial element of good inflation performance, and the empirical evidence to date is consistent with that view (see Alesina and Summers, 1993). Because of this theoretical and empirical evidence, legislation has been introduced around the world that aims at increasing the independence of central banks.

Why do the implications for central bank independence forthcoming from the principal-agent story described above differ so much from what is actually happening in the world? A likely explanation is that this principal-agent story is not the correct view.

Elected leaders As Monetary Authority

Consider an alternative principal-agent problem proposed by Fratianni, von Hagen and Waller (1995). Suppose that voters face an agency problem with elected leaders. Voters want leaders to carry out policies consistent with their objective function in equation 22, but leaders may have incentives to misuse monetary policy for political reasons. For example, elected leaders may follow policies that benefit special interest groups or that further their short-run re-election chances. If unusually high levels of output increase an incumbent's chances of being re-elected, he may try to create surprise inflation to expand output above trend. Furthermore, significant partisanship in the policy process may lead to a redistribution of resources that does not promote the public good. These are all reasons the elected leaders may have an objective function similar to equation 23, if they controlled monetary policy directly.

If elected leaders have an incentive to misuse monetary policy, it is in society's interest to delegate policy to a non-political agent who will enact the policies desired by the general public. This agent would have society's objective function as his own. The problem is: How is this non-political agent chosen? Elections will not work since getting re-elected may be why policy is misused in the first place. The central banker needs to be appointed, but this is typically done by the elected leaders.(6) Thus, elected leaders can use appointment or the threat of non-reappointment to pressure the central bank into implementing policies aimed at helping the incumbent leaders. If the central bank's budget or the central bankers' salaries are under legislative control, then the central bankers can be pressured through budgetary cuts to pursue sub-optimal policies.

In this framework, the central bank would like to do the right thing but its immediate principal--the elected leaders--have objectives that differ from the general public. The elected leaders, not the central bank, need to be made accountable. Accordingly, society benefits by making the central bank as free of political interference as possible, since inflation will be reduced and output will be stabilized optimally. Thus, central bank independence is crucial for good monetary policy; without it, the central bank is merely a veil for political leaders. Anything that makes the central banker's appointment and budget less susceptible to political pressure will lead to better monetary policy.(7) This view of the principal-agent nature of monetary policy has led academic economists to support the movement toward greater central bank independence.

What would be the purpose of central bank performance contracts in this latter version of the principal-agent problem? If the elected leaders are the ones who write and enforce the central bank's performance contract, then they probably will not solve the problem. Clearly, enforcement of the contracts would lack credibility since elected leaders have an incentive to forgive any transgressions the central bank makes (as long as the transgressions benefit the elected leaders).

There is one potential benefit of using performance contracts in this environment. Performance contracts make policy more visible and the goals of the monetary authority more transparent. Presumably, this visibility would lead to better policy actions, since deviations from the socially optimal path would have to be explained publicly at specified **intervals** of time. Individuals who employ political pressure on the central bank would be brought into the public limelight and the personal costs to elected leaders from this attention, we hope, would deter them from putting pressure on the central bank. Furthermore, although it is a blunt instrument, the ballot box may provide enough credibility in the enforcement of the contract such that better macroeconomic performance would be achieved.

CONCLUSIONS

Although theoretically appealing, performance contracts may not be feasible in practice. In fact, political infeasibility may well be the reason we do not observe this type of institutional arrangement in the real world. Nevertheless, the performance contract research we see today could well turn out to be the foundation for the design of central banks in the 21st century. But we'll need to try a few experiments first to see how well they work in practice. New Zealand's recent reforms of its central bank structure seem to be very similar to a performance contract and may well be the test case we need. Evidence to date is sparse, but the reforms appear to have played a role in reducing inflation and inflation expectations.(8)

Future designs of central bank institutions will probably reflect a combination of independence and performance contracts. The result would be highly autonomous central banks that are clearly held accountable to the electorate. What more could we ask for?

REFERENCES

Alesina, Alberto. "Alternative Monetary Regimes: A Review Essay," Journal of Monetary Economics (1988), pp. 175-84.

-- and Roberta Gatti. "Independent Central Banks: Low Inflation at No Cost?" American Economic Review Pope and Proceedings (1995), pp. 196-200.

-- and Lawrence Summers. "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence," Journal of Money, Credit, and Banking (1993), pp. 151-62.

Backus, David, and John Driffill. "Inflation and Reputation," The American Economic Review (1985), pp. 530-38.

Barro, Robert J. "Reputation in a Model of Monetary Policy with Incomplete Information," Journal of Monetary Economics (1986), pp. 1-20.

-- and David B. Gordon. "A Positive Theory of Monetary Policy in a Natural Rate Model," Journal of Political Economy (1983a), pp. 589-610.

-- and -- "Rules, Discretion and Reputation in a Model of Monetary Policy." Journal of Monetary Economics (1983b), pp. 101-22.

Canzoneri, Matthew. "Monetary Policy Games and the Role of Private Information," The American Economic Review (1985), pp. 1056-70.

Faust, Jon. "Whom Can We Trust to Run the Fed: Theoretical Support for the Founders' Views," working paper (1994).

Frotionni, Mike, Juergen von Hagen and Christopher J. Waller, "Central Banking as a Political Principal Agent Problem," Economic Inquiry (forthoming).

Flood, Robert, and Peter Izard. "Monetary Policy Strategies," International Monetary Fund Staff Papers (1989), pp. 612-32.

Garfinkel, Michelle, and Seonghwan Oh. "Strategic Discipline in Monetary Policy with Private Information: Optimal Targeting Horizons," The American Economic Review (1993), pp. 99-117.

Haubich, Joseph, and Joseph Ritter. "Dynamic Commitment and Imperfect Policy Rules," Federal Reserve Bank of St. Louis Working Paper No. 95-015 (November 1995).

Hutchison, Michael. "Central Bank Credibility and Disinflation in New Zealand," Federal Reserve Bank of San Francisco Weekly letter (February 1995).

Kydland, Finn, and Edward Prexott. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," Journal of Political Review (1977), pp. 473-92.

Johmann, Susanne. "Optimal Commitment in Monetary Policy: Credibility Versus Flexibility," The American Economic Review (1992), pp. 273-86.

McCallum, Bennett. "Two Fallacies Concerning Central Bank Independence," The American Economic Review Papers and Proceedings (1995), pp. 207-11.

Persson, Torsten, and Guido Tohellini. "Designing Institutions for Monetary Stability," Carnegie-Rochester Conference Series on Public Policy (1993), pp. 55-83.

Rogoff, Kenneth. "Reputational Constraints on Monetary Policy," Carnegie--Rochester Conference Series on Public Policy (1981), pp. 141-82.

-- "The Optimal Degree of Commitment to on Intermediate Target," Quarterly Journal of Economics (1985), pp. 1169-90.

Spiegel, Mork. "Rules vs. Discretion in New Zealand Monetary Policy," Federal Reserve Bank of San Francisco Weekly Letter (March 1995).

Svensson, Lars. "Inflation Targeting Need Not Increase Output Variability: Comment on Rogoff and Walsh," working paper (1995), Institute for International Economic Studies, Stockholm University.

Thompson, Earl. "What Should Control the Money Supply," The American Economic Review Papers and Proceedings (1981), pp. 356-61.

Waller, Christopher J. "Appointing the Median Voter to a Policy Board," working paper (1995).

-- "A Bargaining Model of Partisan Appointments to the Central Bank," Journal of Monetary Economics (1992).

-- and Carl E. Walsh. "Central Bank Independence, Economic Behavior, and Optimal Term lengths," working paper (1995).

Walsh, Carl E. "Is New Zealand's Reserve Bank Act of 1989 on Optimal Central Bank Contract?" Journal of Money, Credit, and Banking (forthcoming).

-- "Optimal Contracts for Central Bankers," The American Economic Review (1995), pp. 150-67.

"When Should Central Bankers Be Fired?" working paper (September 1994), University of California at Santo Cruz.

1 Recently Haubrich and Ritter (1995) have argued that this comparison between simple rules and discretion is biased in favor of rules, because it assumes that the choice between adopting a simple rule over discretion is a onetime decision. In fact, monetary authority has the option of waiting before committing to a k-percent rule, and this option has **value** that is typically ignored in the Alesina and Johmann analyses. Thus, they argue that discretion is more likely to be preferred than is typically shown.

2 Faust (1994) has shown that the appointment of a central banker who prefers a lower trend inflation rate than the median voter can improve social welfare if the majority of voters are net nominal debt holders. Stabilization issues, however, are not studied in Faust's model.

3 Empirical evidence on this point is mixed. For example, some researchers have shown that greater central bank independence is associated with lower average inflation rates but has no relationship with the variance of GDP. Other work has shown that countries with independent central banks tend to suffer greater output losses during disinflations, which suggests that there is a trade-off between reducing inflation and stabilizing output variability.

4 Peason and Tabellini (1993), working from an early draft of Walsh's paper, extended his approach to a more general framework.

5 Actually, this would imply that the central banker has a different inflation rate target than society's rather than a different output target.

6 The interested reader should see Waller (1992, 1995) for an example of such an appointment process.

7 Waller (1992, 1995), Waller and Walsh (1995) and Alesina and Gatti (1995) show how reducing the degree of political influence in the appointment process can lead to superior macroeconomic outcomes.

8 See Hutchison (1995) and Spiesel (1995).

Christopher J. Waller is an associate professor of economics at Indiana University and served as a visiting scholar at the Federal Reserve Bank of St. Louis when this article was begun. Stephen M. Stohs provided research assistance. The author thanks Joe Ritter, Chris Neely, Bill Gavin, Carl Walsh and Steve Stohs for their comments on earlier drafts.

Special Features: Equations; References
Classification: 9190 (CN=United States); 1120 (CN=Economic policy & planning); 4310 (CN=Regulation); 9130 (CN=Experimental/Theoretical)
Descriptors: Central banks; Public officials; Contracts; Studies; Pay for performance; Monetary policy
Geographic Names: US

36/9/15 (Item 15 from file: 268)
00274329
Evaluating the efficiency of commercial banks: Does our view of what banks do matter?
Wheelock, David C; Wilson, Paul W
Federal Reserve Bank of St. Louis Review , v 77 , n 4 , p 39-43+ , Jul/Aug 1995
CODEN: FRBRDV **Document Type:** Journal Article **ISSN:** 0014-9187 **Journal Code:** FSL **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 06998

Abstract:

Studies have found considerable inefficiency among commercial banks. On average, technical and overall efficiency is higher under the intermediation view of the banking firm than under the production view. Mean allocative efficiency is, however, similar under the extreme versions of each approach. Under the intermediation view, there is somewhat less scale inefficiency and more banks operating on the constant-returns portion of the efficient frontier. Despite the differences in mean measured efficiency across the different conceptions of how banks operate, there is some similarity in the rankings of efficiency scores of individual banks.

Text:

In the past 15 years, the banking industry has faced growing competition from other **financial** service firms and **financial** markets and, at the same time, has undergone substantial deregulation and change. Proponents of further deregulation, such as the removal of barriers to the commingling of commercial and investment banking, argue that such changes would enhance the efficiency and viability of American banks.

The impact of competitive and regulatory changes on banks can be judged by gross measures of performance, such as profitability and failure rates. Economists are also interested in how such changes affect the efficiency with which banks transform resources into various **financial** services. Inefficiency implies that resources are wasted, that is, that firms are producing less than the feasible level of output from the resources employed, or are using relatively costly combinations of resources to produce a particular mix of products or services. Thus, a goal of policymakers, as well as stockholders and managers, is to devise policies that improve the efficiency of commercial banks.

Unfortunately, economists do not agree upon the appropriate methodology for measuring the efficiency of banks. Several estimation techniques have been proposed, each with advantages and disadvantages. The problem is complicated by the myriad of different services that commercial banks perform. Researchers deal with complex issues in measuring bank

production: Is a deposit an input to the production process, or an output! Should outputs be measured in terms of the number of a bank's accounts, the number of transactions it processes or the dollar amounts of its loans or deposits? Perhaps not surprisingly, estimates of commercial bank inefficiency vary considerably across studies that use different techniques, conceptions of bank production and data samples.

This article investigates the sensitivity of efficiency measures to broadly different conceptions of how banks operate. We use a single-estimation technique and a common pool of banks to compare efficiency measures based on alternative views of bank production. We find substantial differences in mean efficiency across models and low, though statistically significant, correspondence in the rankings of banks by efficiency scores across models.

First, we discuss why measuring commercial bank efficiency is useful, some alternative measures of efficiency and techniques for estimating efficiency. A description of the approach we take, our data and our results follow.

WHY DO WE CARE ABOUT THE EFFICIENCY OF COMMERCIAL BANKS?

The performance of firms is often described in terms of their efficiency. The measured efficiency of a production unit (a firm or plant) is generically interpreted as the difference between its observed input and output levels and the corresponding optimal values. An output-oriented measure of efficiency compares observed output with the maximum output possible for given input levels. Alternatively, an input-oriented efficiency measure compares the observed level of inputs with the minimum input that could produce the observed level of output. These are measures of technical efficiency, and as such ignore the behavioral goals of the firm.

Measures of allocative efficiency compare the observed mix of inputs or outputs with the optimal mix that would minimize cost, maximize profit or obtain any other behavioral goal. Allocative efficiency can be combined with technical efficiency to measure overall efficiency. In addition, measures of technical efficiency can be used to construct measures of scale efficiency, which involve comparison of observed and optimal scale, or size, of the firm. One can also measure scope efficiency, which involves comparison of the cost of producing the observed mix of outputs in a single firm with the costs that would prevail if each output was produced in a separate firm. Researchers have found that banks suffer more from technical inefficiency than from scale or scope inefficiency (for example, Berger and Humphrey, 1991).

The efficiency of commercial banks is important for at least two reasons. First, efficiency measures are indicators of success, by which the performance of individual banks, and the industry as a whole, can be gauged. Banks face growing competition, both from other banks and from firms and markets outside the industry (see Wheelock, 1993), and presumably banks will be more successful in maintaining their business if they operate efficiently. Berger and Humphrey (1992) find that during the 1980s high-cost banks experienced higher rates of failure than more efficient banks. Similarly, in a study of bank failures during the 1920s, Wheelock and Wilson (1995) find that the less technically efficient a bank was, the greater its likelihood of failure.

A second reason to investigate the efficiency of commercial banks is the potential impact of government policies on efficiency. One might gauge the impact of a regulatory change by measuring its effect on commercial bank efficiency, or examine efficiency among banks in different states to measure the effect of differences in branching restrictions or other regulations. Recent proposals to end the Glass-Steagall separation of commercial and investment banking stem in part from a view that broader powers could enhance the efficiency of banks and other **financial** institutions. Obviously, this change could enhance the scope efficiency of banks if there are complementarities in the production of commercial and investment banking services. Conceivably, such change could also improve

scale or overall efficiency. Improved efficiency is also one argument made in support of interstate branching and, indeed, Grabowski, Rangan and Rezvanian (199) find that branch banking organizations are more efficient than multiple-office bank holding companies.

Other studies have considered whether bank mergers enhance efficiency. Using different approaches, Rhodes (1993) finds that mergers have not generally improved efficiency, though Fixler and Zieschang (1993) conclude the opposite. Shaffer (1993), on the other hand, evaluates potential mergers and concludes that they could significantly reduce inefficiency for many banks of less than $10 billion of assets.

The impact of ownership or management structure on efficiency has also been studied. Pi and Timme (1993), for example, find that banks whose chief executive officer also serves as board chairman are less efficient than other banks, and Mester (1993) shows that mutual savings and loan associations are more efficient than stock S&Ls.

MEASURING COMMERCIAL BANK EFFICIENCY

The efficiency of commercial banks has been studied using a variety of techniques and samples, and, as noted above, has been used to address numerous policy issues. Recent studies typically use techniques that accommodate the multiple outputs of banks and measure the efficiency of individual banks relative to a standard set by peer institutions. Readers interested in a survey of this research can refer to Berger, Hunter and Timme (1993).

To date, no technique for measuring efficiency has been generally accepted and different methodologies appear to generate considerable differences in measured efficiency, even when common bank samples are used. Variants of four techniques are common in the literature. The "stochastic cost frontier" approach is an econometric methodology in which deviations of a firm's actual cost from predicted cost are presumed to be due to **random** error and inefficiency, each of which is assumed to have a particular statistical distribution (usually the normal distribution for the **random** error and a half-normal for inefficiency). The "thick frontier" approach is a variant in which deviations from predicted cost within the lowest average cost quartile of banks are assumed due to **random** error, and the differences between the predicted costs of banks in the highest and lowest quartiles are assumed to be due to inefficiency. The "distribution-free" approach is applicable when data for more than one year are available. It assumes that inefficiency is stable over time, while **random** errors average out over time. That is, a bank's inefficiency for a span of years is taken to be the mean of its measured inefficiency across all years within the period. Finally, "Data Envelopment **Analysis**" (DEA) is a non-parametric methodology in which linear programming is used to measure the distance of individual banks from the efficient, or "best-practice," frontier. All deviations from the efficient frontier are assumed to be due to inefficiency.

Researchers have found that estimates of inefficiency are sensitive to the choice of technique. Ferrier and Lovell (1990), for example, apply the stochastic cost frontier and DEA techniques to a common sample of banks and arrive at different estimates of inefficiency. Berger (1993) finds substantial differences in measured efficiency from two variants of the distribution-free approach.

A second reason why different studies of commercial bank efficiency often reach seemingly contradictory findings might stem from differences in how a banking firm is modeled. Regardless of which of the four measurement techniques is used, the researcher must specify a list of inputs and outputs. The question, "What do banks produce?" is not simple to answer. Banks provide a variety of services, from loans and deposit accounts to trust services, safe deposit box rentals, mutual fund sales and foreign exchange transactions. Moreover, changes in regulation, technology and customer demands have caused the types of services that banks perform to change over time. For example, banks now provide a variety of securities-related services, such as underwriting and mutual fund sales,

which regulators forbid a few years ago. To tractably measure efficiency, researchers are forced to begin with simplified models of the banking firm. Unreliable estimates of efficiency can stem from the use of models that omit key features of bank production.

Some researchers view banks as producers of loans and deposit accounts, and measure output by either the number of transactions or accounts serviced. This view is referred to as the "production" approach. Others argue that a bank's output should be measured in terms of the dollar volume of loans or deposits it provides, a view known as the "intermediation" approach. Most studies of inefficiency use the intermediation approach, in part because the necessary data are more readily obtained. We are aware of only one recent study taking the production approach (Ferrier and Lovell, 1990), though in the 1970s and early 1980s such studies were more common (see Gilbert, 1984). The production approach focuses on operating costs and ignores interest expense. The intermediation approach, on the other hand, includes both operating and interest expenses, and hence may be of more interest for studying the viability of banks (see Berger, Hanweck and Humphrey, 1987; or Ferrier and Lovell, 1990). For **analysis** of the operating efficiency of banks, however, the production approach may be of interest.

Among those who use the intermediation approach are researchers who hold the view that banks produce various loans and other investments from deposits, other funding sources, labor and materials. This "asset" approach has been criticized because it ignores the fact that banks expend considerable resources supplying transactions and savings deposits (Berger and Humphrey, 1992).

Some researchers apply empirical criteria to determine what services to consider as bank outputs and what to consider as inputs. Berger and Humphrey (1992), for example, classify activities for which banks create high added **value**, such as loans, demand deposits and time and savings deposits as important outputs, with labor, physical capital and purchased funds classified as inputs. Alternatively, Aly, Grabowski, Pasurka and Rangan (1990), Hancock (1991) and Fixler and Zieschang (1993) adopt a "user-cost" framework, whereby a bank asset is classified as an output if the **financial** return on the asset exceeds the opportunity cost of the investment, and a liability is classified as an output if the **financial** cost of the liability is less than its opportunity cost. Even though their details differ, the two approaches empirically tend to suggest similar classifications of inputs and outputs. The main exception is classification of demand deposits as an output in most user-cost studies, and as both an input and an output when the **value**-added approach is used (see Berger and Humphrey, 1992, for more detail).

Table 1 summarizes six recent studies of commercial bank production efficiency. (Table 1 omitted) Although representative, this list is far from exhaustive. These studies employ a variety of estimation techniques and include a variety of different inputs and outputs in modeling the banking firm. The studies typically report inefficiency measures by bank-size grouping and for more than one type of inefficiency, though for brevity we report just the mean overall inefficiency. The reported percentages indicate the extent to which the average bank overused inputs to produce a given level of output. Thus, the 35 percent inefficiency found by Aly and others (1990) indicates that the average bank could have produced the same level of output with just 65 percent of the input levels actually used. Measured inefficiency clearly varies with estimation technique, model specification and, probably, the sample of banks used by the researcher.

In the remainder of this article, we investigate the extent to which measures of efficiency and the rankings of individual banks depend on whether the intermediation approach or production approach is employed. Because we are interested in the impact of the approach taken on measured efficiency, we use a single technique--DEA--applied to a common pool of

banks. Our findings might, of course, be different if we used another technique or sample, but the purpose of this article is to investigate how sensitive efficiency measures are to the model of bank production employed.

METHODOLOGY

We trace our measures of efficiency to the work of Debreu (1951) and Farrell (1957). Boles (1966) was one of the first to use linear programming methods to measure efficiency in production using their ideas. Other extensions have collectively come to be named Data Envelopment **Analysis** (DEA), a term coined by Charnes, Cooper and Rhodes (1978). Lovell (1993) summarizes this literature.

Details about the efficiency measures used in this article are contained in the shaded insert on page 6. The essential ideas, however, are illustrated in Figure 1, which considers the case of a sample of firms producing a single output from two inputs, x sub 1 and x sub 2 . (Figure 1 omitted) Suppose firms A, B and C each produce a given level of output; A and B lie on the production frontier XX', while C lies in the interior of the production set. The frontier XX' is the set of all combinations of inputs which can produce the same level of output, and where the reduction of at least one input necessarily causes output to fall. Hence, firms A and B are regarded as efficient, whereas firm C is regarded as inefficient. Inefficient firms such as C may lie in the interior of the production set due to imperfect information, managerial incompetence or perhaps other reasons. For firm C, input weak technical efficiency (IWE) is defined as the ratio of distances OC'/OC in Figure 1. By reducing the input quantities used by firm C by this amount, the firm could move to point C' and would be considered efficient in the IWE sense.

Next, we define input overall efficiency (IOE). In terms of Figure 1, the isocost line is given by PP'. For firm C, the IOE score is given by the ratio of distances OC"/OC. Although the point C" lies outside the production set boundary, and hence is not feasible, input costs at C" are the same as at B, which is a feasible point. Hence, if firm C were to become efficient in the IOE sense, its input mix would have to be altered; the IOE score, however, can be obtained by considering the hypothetical proportionate reductions of inputs represented by point C".

In terms of Figure 1, allocative efficiency for firm C is given by the ratio of distances OC"/OC'. Allocative inefficiency arises from using a combination of inputs that does not minimize total cost, as opposed to technical inefficiency, which is a proportionate overuse of all inputs.

Finally, we can determine scale efficiency by comparing IWE computed under the assumption that the firm is operating at constant returns-to-scale with IWE obtained previously. A score of unity implies that the firm is operating under constant returns. While a score other than 1 does not translate easily into a specific percentage deviation from constant returns, the scores are useful for ranking firms by the extent of their inefficiency.

Each of the efficiency scores described above measures efficiency in an input orientation; efficiency is measured by holding output fixed and determining the maximum feasible reduction in inputs. Efficiency can also be measured by holding inputs fixed and determining the maximum feasible expansion of outputs. Since the efficiency measures we use do not imply underlying assumptions regarding the behavior of firms, the choice between input and output orientations is somewhat arbitrary; one might compute both types of efficiency measures to get more information than can be obtained from either the input or output orientations alone. Note that both IWE and IOE are radial measures of efficiency, that is, in each case efficiency is measured along a ray emanating from the origin and passing through the firm in input-output space. Consequently, the efficiency scores are independent of the units of measurement used for both inputs and outputs, which is advantageous since units of measurement may always be defined arbitrarily. Faere and others (1985) observe that some DEA formulations do not share this property.

EMPIRICAL IMPLEMENTATION

For our empirical **analysis** of commercial bank efficiency, we use a sample of banks participating in the Federal Reserve System's Functional Cost **Analysis** (FCA) program for 1993. Participants in this program supply information about their operations and costs which are not generally available for banks, and which are necessary to measure efficiency using the production approach. After eliminating observations with missing values and observations for depository institutions other than commercial banks, data for 269 banks remain.

Because participation in the FCA program is voluntary, the banks in our sample may not be representative of the industry as a whole. For example, whereas the average total assets at the end of 1993 for FCA program banks was $163.6 million, with a range from $8.0 million to $2,602.8 million, average total assets were $300.7 million, with a range from $1.0 million to $108,223.0 million, for all U.S. commercial banks (as reported in the Federal Deposit Insurance Corporation Reports of Condition, that is, the "Call Reports"). The average return on assets of 1.15 percent for the banks in our sample, however, was approximately the same as the average for all banks (1.12 percent). Nevertheless, because our sample of banks is not **random**, the efficiency measures calculated here should not be interpreted as reflecting the efficiency of commercial banks in general.

For the production approach to modeling bank activities, we construct variables using definitions from Ferrier and Lovell (1990):

Outputs:

y sub 1 = number of demand deposit accounts

y sub 2 = number of time deposit accounts

y sub 3 = number of real estate loans

y sub 4 = number of installment loans

y sub 5 = number of commercial loans

Inputs:

x sub 1 = number of employees

x sub 2 = occupancy costs and expenditure on furniture and equipment

x sub 3 = expenditure on materials

Input prices:

w sub 1 = total expenditure on salaries and fringe benefits/x sub 1

w sub 2 = x sub 2 /level of deposits

w sub 3 = x sub 3 /level of deposits

For the intermediation approach to bank production, we construct variables using definitions from Kaparakis, Miller and Noulas (1994):

Outputs:

v sub 1 = loans to individuals for household, family, and other personal expenses

v sub 2 = real estate loans

v sub 3 = commercial and industrial loans

v sup 4 = federal funds sold, securities purchased under agreements to resell, plus total securities held in trading accounts

Inputs:

u sub 1 = interest-bearing deposits except certificates of deposit greater than $100,000

u sub 2 = purchased funds (certificates of deposit greater than $100,000, federal funds purchased, and securities sold plus demand notes) and other borrowed money

u sub 3 = number of employees

u sub 4 = premises and fixed assets

Input prices:

p sub 1 = average interest cost per dollar of u sub 1

p sub 2 = average interest cost per dollar of u sub 2

p sub 3 = average annual wage per employee

p sub 4 = average cost of premises and fixed assets.

In addition, Kaparakis and others also define quasi-fixed input, non-interest bearing deposits, for which there is no corresponding price. Other studies adopting the intermediation approach have ignored this item,

as we do in the results reported below. Including non-interest bearing deposits as a fifth input when measuring technical or scale efficiency seems to have little effect on the results.

To form a specification midway between the production approach represented by the Ferrier and Lovell (1990) specification and the intermediation approach represented by the Kaparakis and others (1994) specification, we define y' sub 1 ,...,y' sub 5 as the dollar amount of each account or loan corresponding to y sub 1 ,...,y sub 5 , respectively. Because outputs are now measured in dollar amounts, this model is best classified as representing the intermediation approach, even though the choice of variables is based on Ferrier and Lovell (1990). In addition, we define an alternative price system, w' sub 1 , w' sub 2 , w' sub 3 , for the Ferrier and Lovell specification, where w' sub 1 = w sub 1 , and w' sub 2 and w' sub 3 are computed similarly to w sub 2 and w sub 3 except that level of deposits is replaced by the number of time and demand deposits. This seems to us to make the mapping of inputs and outputs under the production approach more consistent. We report summary statistics for each of the variables in Table 2. (Table 2 omitted)

RESULTS OF EFFICIENCY MEASUREMENT

We compute the various efficiency measures for the five models summarized below:

(Models omitted)

Models 1 and 2 correspond to the Ferrier and Lovell (1990) specification, with alternative price definitions. Models 3 and 4 provide a bridge to the intermediation approach by replacing the number of accounts and loans in the output variables with dollar amounts. Model 5 is the Kaparakis and others (1994) specification.

Table 3 presents the mean scores for each type of efficiency described in the preceding section. (Table 3 omitted) Note that since the same inputs and outputs are used in models 1 and 2, and models 3 and 4, the technical and scale efficiency scores are the same for these models. For each efficiency measure, Table 3 also shows the standard deviation of the scores across the 269 banks in the sample, the number of banks having an efficiency score of unity (labeled "Number Efficient"), that is, the number of banks operating on the efficient frontier, as well as 90 and 95 percent confidence **intervals** for the mean. Given the large number of banks with efficiency scores of unity, particularly in the case of technical efficiency, and since all of the efficiency scores are defined to lie between zero and 1, the underlying distributions of the individual efficiency scores are clearly non-normal. Results from Atkinson and Wilson (1995), however, suggest that our sample size of 269 is easily large enough for us to rely on the asymptotic normality of the sample means implied by the central limit theorem, and thus to compute confidence **intervals** based on a normal distribution.

In several cases, the confidence **intervals** for the means reported in Table 3 overlap. We test for significant differences among the means of each efficiency measure across different models. At the 0.05 significance level, we are unable to reject the null hypothesis of equivalent means in the following cases: (1) scale efficiency for models 3, 4 and 5; (2) allocative efficiency for models 1 and 5, and models 3 and 5 (we do reject the null hypothesis when comparing allocative efficiency among models 1 and 3); and (3) overall efficiency for models 1 and 4. In all other instances, we reject the null hypothesis of no difference. Even the seemingly innocuous modification of redefining the input prices between models 1 and 2, and between 3 and 4 has a large effect on mean allocative and overall efficiency. Note also that for the most extreme comparison, models 1 (the production view) and 5 (the intermediation view), we reject the null hypothesis of equal levels of technical and overall efficiency. This suggests that, at least for this sample of banks, average efficiency does depend on the view of bank production assumed by the researcher. We find that average technical and overall efficiency is higher under the intermediation approach (model 5) than under the production approach (model

1). Our finding for overall inefficiency of 37 percent using model 5 is similar to what Aly and others (1990) found for their sample, though substantially greater than what Kaparakis and others (1994) found for theirs (see Table 1).

It is possible to determine whether a particular bank lies on the increasing (IRS), constant (CRS) or decreasing (DRS) returns portion of the technology. Table 4 shows the results of this **analysis**, considering only banks that were found to be technically efficient.(1) (Table 4 omitted) Thus, for example, 16 banks, or 38.1 percent of all technically efficient banks, operated on the constant-returns portion of the technology under models 1 and 2.

We test the null hypothesis of no association among the rows and columns of the matrix represented by Table 4 using Pearson's chi-square test, the likelihood ratio chi-square statistic, and Fisher's exact test.(2) For the entire matrix, all three tests reject the null hypothesis of no differences in the proportions in each row and column. However, when we perform pairwise tests by deleting individual rows from Table 4, we fail to reject the null hypothesis of no difference for models 3, 4 and 5, and for models 1, 2 and 3, 4. Each of the three tests fail to reject at the 90 percent level.

In the case of models 1, 2 and 5, we reject the null hypothesis of no difference in the proportions at greater than 99 percent. Thus, while we find evidence of similarity in terms of returns-to-scale when comparing models 3, 4 with either models 1, 2 or 5, models 1, 2 and 5 appear different in terms of returns-to-scale. More banks appear to be operating under constant returns-to-scale when the intermediation approach is taken (model 5) than when the production approach is used (model 1). Since returns-to-scale at a given location on the production frontier depend upon the shape of the variable-returns technology, these results indicate that the technology implicitly estimated by models 3, 4 is similar to the technologies implied by models 1, 2 and 5, which in turn are significantly different. This is consistent with our view of models 3, 4 as a bridge between the production approach represented by models 1, 2 and the intermediation approach represented by model 5. The result also suggests that differences between the two approaches might be due not only to use of number of accounts and loans versus dollar amounts, but also to the treatment of time deposits as an output or an input.

In addition to comparing mean efficiency scores, we use the Wilcoxon matched-pairs signed-ranks test, a sign test for equality of medians and Kendall's tau-statistic to further examine the similarity of efficiency scores across the five models. We report the results of these tests in Table 5.(3) (Table 5 omitted)

The Wilcoxon matched-pairs signed-ranks test analyzes the equality of distributions without making assumptions regarding the form the distributions might take. The values shown in parentheses in the second column of Table 5 give the two-tailed normal probabilities associated with the test statistic. Hence, we fail to reject the null hypothesis of identical distributions when comparing scale efficiency scores from models 3, 4 and 5, when comparing allocative efficiency scores from models 1 and 5, and from models 3 and 5, and when comparing overall efficiency scores from models 1 and 4. In all other cases, we reject the null hypothesis. It appears that, for the most part, the distributions of the various efficiency scores do vary across models.

The sign test for equivalence of medians yields a two-tailed binomial probability, which we also report in Table 5. In only two instances do we not reject the null hypothesis of equal medians: when comparing scale efficiency scores from models 3, 4 and 5, and when comparing allocative efficiency scores from models 3 and 5. These results are consistent with our finding that, in most cases, average efficiency varies across models.

Finally, rather than comparing the distributions of efficiency scores from different models, we use the scores to rank banks in terms of their estimated efficiency. Kendall's tau-statistic measures the correlation

among the ranks of banks from two models and provides a statistical test of the null hypothesis of no association between two sets of rankings. The statistic is approximately normally distributed, with zero expected **value** and with variance

(Equation omitted)

where N gives the number of observations. By definition, the statistic lies between -1 and +1, taking a **value** of +1 if rankings are in complete agreement, or -1 if the ranks are completely reversed.

Kendall and Gibbons (1990) suggest that the tau-statistic may also be viewed as a measure of concordance. Any two pairs of ranks (u sub i , v sub i) and (u sub j , v sub), i, j = 1,...,N, i -=, are defined as concordant if v sub i < v sub when u sub i < u sub j or v sub i > v sub j when u sub i > u sub j . Similarly, they are defined as discordant if v sub i < v sub j when u sub i > v sub j or v sub i > v sub j when u sub i > u sub j . The total number of pairs is N(N-1)/2, and tau can be shown to be equivalent to the proportion of concordant pairs minus the proportion of discordant pairs.

The last column of Table 5 gives the tau-statistic for the various pairs of models for each measure of efficiency, along with significance levels as shown in parentheses. (Table 5 omitted) We fail to reject the null hypothesis of no association among ranks in only five instances when comparing allocative efficiency scores. (In particular, we fail to reject the null hypothesis for the following pairs of models: 1, 2; 2, 3; 2, 5; 3, 4; and 3, 5, and in only one case when comparing overall efficiency scores--for models 2 and 5). In all other cases, we reject the null hypothesis of no correlation.

Note, however, that when we reject the null hypothesis of no association, the tau-statistic is usually rather small in absolute terms; the largest **value** of the statistic shown in Table 5 is 0.5874 (in the case of overall efficiency for models 1 and 2). As is typical in classical hypothesis testing, rejection of the null hypothesis does not necessarily imply acceptance of an alternative hypothesis. That is, our statistical test may reject the hypothesis that the rankings are not associated, but that does not necessarily imply that the rankings are associated-the test is simply not that powerful. Figure 2 plots the rankings of overall efficiency scores for model 1 against those for model 5. Note that the **value** of Kendall's tau-statistic for this comparison is significantly different from zero at the 0.0012 level, indicating that the two sets of rankings are not discordant. The low **value** of the test statistic (0.1329), however, suggests that neither are they concordant.

Our results based on the Wilcoxon matched-pairs signed-ranks test and the sign test for equivalence of medians are consistent with our observations on the differences of mean efficiency scores across the models discussed earlier. Taken together, our results indicate that different model specifications are likely to produce different measures of the level of inefficiency among a sample of banks, but not necessarily dissimilar rankings of individual banks in terms of measured efficiency. For our data, the rankings are similar enough to reject the null hypothesis of no association, but in many cases are far from being in complete concordance. Concordance is relatively high for the technical and scale efficiency measures, which do not rely on price data. The introduction of price data to measure allocative and overall efficiency might also introduce more sources of noise or error.

CONCLUSION

Like other studies of commercial bank efficiency, we find considerable inefficiency among banks in our sample. Other studies have found substantial variation in efficiency measures in applying different estimation techniques to a common pool of banks. We find that measured efficiency also depends on the researcher's conception of what banks do. In this article, we measure various types of production efficiency under two very different views of banking. We find that, on average, technical and

overall efficiency is higher under the intermediation view of the banking firm than under the production view. Mean allocative efficiency is, however, similar under the extreme versions of each approach. Under the intermediation view, we also find somewhat less scale inefficiency and more banks operating on the constant-returns portion of the efficient frontier. Despite the differences in mean measured efficiency across the different conceptions of how banks operate, however, we find some similarity in the rankings of efficiency scores of individual banks. Further research will, of course, be necessary to determine how sensitive these findings are to the particular dataset and estimation techniques employed in this article.

A MATHEMATICAL DESCRIPTION OF EFFICIENCY MEASUREMENT

We use measures of efficiency discussed by Faere, Grosskopf and Lovell (1985). First, we compute the input weak technical efficiency (IWE) score for the ith firm in a sample by solving the linear programming problem:

min W sub i

subject to Xq sub i <= W sub i x sub i

(Equation 1 omitted)

where n firms produce s outputs using m inputs, q sub i is a (N X 1) vector of weights to be computed for the ith firm, 0 < W sub i <= 1 is a scalar, x sub i is a (m X 1) vector of inputs for the kth firm, y sub i is a (s X 1) vector of outputs for the kth firm, X = [x sub i ,...,x sub N] is a (m X N) matrix of observed inputs, Y= [y sub 1 ,...,y sub N] is a (s X N) matrix of observed outputs and (character omitted) is a (1 X N) vector of ones.

The minimand W sub i in equation 1 measures the input weak efficiency of the ith firm. The inequality constraints in equation (character omitted) define a reference technology with strong disposability of inputs; constraining the weights in q to sum to unity allows the reference technology to exhibit variable returns to scale. For the ith firm, W sub i gives the proportion by which inputs can be reduced to move the firm from the interior of the production set onto the piecewise-linear boundary of the production set corresponding to the reference technology in 1.

Next, we compute input overall efficiency (IOE) score O sub i for the ith firm by first solving the linear program:

min p sub i x(characters omitted) x(characters omitted)

subject to (Equation omitted)

(Equation 2 omitted)

where X, Y, x sub i and y sub i are defined as in equation 1, p sub i is a (1 X m) vector of input prices, and x(characters omitted) is an (m X 1) vector of efficient inputs to be computed. The IOE score may be defined as

(Equation 3 omitted)

The constraints in equation 2 are similar to those in 1. The same reference technology is defined by the constraints in 2, but instead of proportionately reducing inputs until the ith firm lies on the reference technology, inputs are further reduced proportionately until the firm lies on the isocost plane tangent to the production set boundary.

An allocative efficiency score, A sub i , may be defined by dividing the IOE score by the IWE score:

(4)

A sub i == O sub i /W sub i .

The efficiency scores obtained from 1 measure technical efficiency as the distance to the relevant isoquant, but do not consider where the firm is situated along the variable- returns production frontier. To measure scale efficiency, equation 1 must be recomputed for each firm, first assuming constant returns to scale by removing the restriction Cq sub i =1, and then assuming non-increasing returns-to-scale by imposing the restriction Cq sub i <= 1. In the case of IWE, this produces efficiency scores W(characters omitted) and W(characters omitted) respectively, for the ith firm. The scale efficiency score corresponding to 1 is then defined as

(Equation 5 omitted)

Clearly, 0 < S sub i <= 1 since W(characters omitted) <= W(characters omitted) <= W sub i . If S sub i = 1, then the ith firm is scale-efficient, that is, the firm is operating at the point of constant returns on the production frontier. If S sub i < 1, then the firm is scale-inefficient due to either decreasing returns if W(characters omitted) = W sub i , or increasing returns if W(characters omitted) < W sub i .

1 Since we are using on input orientation, we could also examine whether inefficient banks would lie on the increasing-, constant-decreasing-returns portion of the technology if inputs were proportionately contracted to move the bank to the frontier. However, since the path a bank might actually take to reach the frontier if the sources of inefficiency were removed depends upon behavioral goals, we ignore technically inefficient banks here.

2 Details on the computations may be found in the State Reference Manual Release 3.1, Stata Corporation (1993).

3 See Snedecor and Cochran (1989) and Kendall and Gibbons (1990) for a discussion of these tests. Computational details are given in the Stata Reference Manual: Release 3.1, Stata Corporation (1993).

REFERENCES

Aly, Hassan Y., Richard Grabowski, Carl Pasurka and Nanda Rangan. "Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation," Review of Economics and Statistics (May 1990),pp. 211-18.

Atkinson, Scott, and Paul W. Wilson. "Comparing Mean Efficiency and Productivity Scores Across Small Samples: A Bootstrap Methodology," Journal of Productivity **Analysis** (forthcoming).

Berger, Allen N. "Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," Journal of Productivity **Analysis** (1993), pp. 261-92.

--, G. A. Hanweck and D. B. Humphrey. "Competitive Viability in Banking: Scale, Scope, and Product Mix Economies," Journal of Monetary Economics (1987), pp. 501-20.

-- and David B. Humphrey. "Measurement and Efficiency Issues in Commercial Banking," in Zvi Griliches, ed., Output Measurement in the Services Sector. University of Chicago Press, 1992, pp. 245-79.

-- and --. "The Dominance of Inefficiencies Over Scale and Product Mix Economies in Banking," Journal of Monetary Economics (August 1991), pp. 117-48.

-- William C. Hunter and Stephen G. Timme. "The Efficiency of **Financial** Institutions: A Review and Preview of Research, Past, Present and Future," Journal of Banking and **Finance** (April 1993), pp. 221-49.

Boles, J. "Efficiency Squared: Efficient Computations of Efficiency Indexes," Western Form Economics Association Proceedings (1966), pp. 137-42.

Charnes, A., W. W. Cooper and E. Rhodes. "Measuring the Efficiency of Decision Making Units," European Journal of Operational Research (1978), pp. 429-44.

Debreu, Gerard. "The Coefficient of Resource Utilization," Econometrica (1951), pp. 273-92.

Elyasiani, Elyas, and Seyed M. Mehdian. "A Nonparametric Approach to Measurement of Efficiency and Technological Change: The Case of Large U.S. Commercial Banks," Journal of **Financial** Services Research (July 1990), pp. 157-68.

Faere, Rolf, Shawna Grosskopf and C. A. Knox Lovell. The Measurement of Efficiency of Production. Kluwer-Nijoff Publishing, 1985.

Farrell, M. J. "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society, Series A (1951), pp. 253-81.

Ferrier, Gary D., and C. A. Knox Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," Journal of Econometrics (October/November 1990), pp. 229-45.

Fixler, Dennis J., and Kimberly D. Zieschang. "An Index Number Approach to Measuring Bank Efficiency: An Application to Mergers," Journal of Banking and **Finance** (April 1993), pp. 437-50.

Gilbert, R. Alton. "Bank Market Structure and Competition: A Survey," Journal of Money, Credit and Banking (November 1984), pp. 617-45.

Grabowski, Richard, Nanda Rangan and Rasoul Rezvanian. "Organizational Farms in Banking: An Empirical Investigation of Cost Efficiency," Journal of Banking and **Finance** (April 1993), pp. 531-38.

Hancock, Diana. A Theory of Production for the **Financial** Firm. Kluwer Academic Publishers, 1991.

Hunter, William C., and Stephen G. Timme. "Core Deposits and Physical Capitol: A Reexamination of Bank Scale Economies and Efficiency with Quasi-Fixed Inputs," Journal of Money, Credit and Banking (February 1995),pp. 165-85.

Kaparakis, Emmanuel I., Stephen M. Miller and Athanasios G. Noulas. "Short-run Cost Inefficiency of Commercial Banks: A Flexible Stochastic Frontier Approach," Journal of Money, Credit and Banking (November 1994), pp. 875-93.

Kendall, M., and J. D. Gibbons. Rank Correlation Methods. Oxford University Press, 1990.

Lovell, C. A. Knox. "Production Frontiers and Productive Efficiency," in Hal Fried, C. A. Knox Lovell and Shelton S. Schmidt, eds., The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, 1993, pp. 3-67.

Mester, Loretta J. "Efficiency in the Savings and Loan Industry," Journal of Banking and **Finance** (April 1993), pp. 261-86.

Pi, Lynn, and Stephen G. Timme. "Corporate Control and Bank Efficiency," Journal of Banking and **Finance** (April 1993), pp. 515-30.

Rhoades, Stephen A. "Efficiency Effects of Horizontal (In-Market) Bank Mergers," Journal of Banking and **Finance** (April 1993), pp. 411-22.

Shaffer, Sherrill. "Can Megamergers Improve Bank Efficiency?" Journal of Banking and **Finance** (April 1993), pp. 423-36.

Snedecor, G. W., and W. G. Cochran. Statistical Methods. Iowa State University Press, 1989.

Stata Reference Manual: Release 3.1. State Corporation, 1993.

Wheelock, David C. "Is the Banking Industry in Decline? Recent Trends and Future Prospects from a **Historical** Perspective," this Review (September/October 1993), pp. 3-22.

-- and Paul W. Wilson. "Explaining Bank Failures: Deposit Insurance, Regulation and Efficiency," Review of Economics and Statistics (forthcoming).

David C. Wheelock is a senior economist at the Federal Reserve Bank of St. Louis. Paul W. Wilson is an associate professor of economics at the University of Texas at Austin. Heidi L. Beyer provided research assistance.

Special Features: Charts; Graphs; References
Classification: 9190 (CN=United States); 9130 (CN=Experimental/Theoretical); 8110 (CN=Commercial banking)
Descriptors: Studies; Commercial banks; Efficiency; Problems
Geographic Names: US

36/9/16 (Item 16 from file: 268)
00270991
Distinguishing theories of the monetary transmission mechanism
Cecchetti, Stephen G
Federal Reserve Bank of St. Louis Review , v 77 , n 3 , p 83-97 , May/Jun 1995
CODEN: FRBRDV **Document Type:** Journal Article **ISSN:** 0014-9187 **Journal Code:** FSL **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 08833

Abstract:

A study examines how one might determine whether the cross-sectional effects of monetary policy are quantitatively important. The goal is to provide a critical evaluation of the major contributions to the literature thus far. It is concluded that the myriad studies have succeeded in establishing the empirical importance of credit market imperfections. This means that monetary policy shifts have an important distributional aspect that cannot be addressed within the traditional money view. It is the smaller and faster growing firms that bear a disproportionate share of the burden imposed by a recession. Since these are likely to be firms with highly profitable investment opportunities, this has important implications for social welfare.

Text:

Traditional studies of monetary policy's impact on the real economy have focused on its aggregate effects. Beginning with Friedman and Schwartz (1963), modern empirical research in monetary economics emphasizes the ability of policy to stabilize the macroeconomy. But casual observation suggests that business cycles have distributional implications as well. One way of casting the debate over the relative importance of different channels of monetary policy transmission is to ask if these distributional effects are sufficiently important to warrant close scrutiny.

The point can be understood clearly by analogy with business cycle research more generally. If recessions were characterized by a proportionate reduction of income across the entire employed population--for example, everyone worked 39 rather than 40 hours per week for a few quarters--then economists would pay substantially less attention to cycles. It is the allocation of the burden or benefit of fluctuations, with some individuals facing much larger costs than others, that is of concern. There are two ways for an economist to address this problem. The first is to attempt to stabilize the aggregate economy, the traditional focus of policy-oriented macroeconomics. The second is to ask why the market does not provide some form of insurance.

The recent debate over the nature of the monetary transmission mechanism can be thought of in similar terms. According to the original textbook IS-LM view of money, changes in policy are important only insofar as they affect aggregate outcomes. Only the fluctuation in total investment is important since policies only affect the required rate of return on new investment projects, and so it is only the least profitable projects (economy-wide) that are no longer funded. But since the most profitable projects continue to be undertaken, there are no direct efficiency losses associated with the distributional aspects of the policy-induced interest rate increase.

In contrast, the "lending" view focuses on the distributional consequences of monetary policy actions. By emphasizing a combination of capital market imperfections and **portfolio**lt; balance effects based on imperfect asset substitutability, this alternative theory suggests the possibility that the policy's incidence may differ substantially across agents in the economy. Furthermore, the policy's impact has to do with characteristics of the individuals that are unrelated to the inherent creditworthiness of the investment projects. An entrepreneur may be deemed unworthy of credit simply because of a currently low net worth, regardless of the social return to the project being proposed. It is important to understand whether the investment declines created by monetary policy shifts have these repercussions.(1)

In this essay, I examine how one might determine whether the cross-sectional effects of monetary policy are quantitatively important. My goal is to provide a critical evaluation of the major contributions to the literature thus far. The discussion proceeds in three steps. I start in the first section with a description of a general framework that encompasses all views of the transmission mechanism as special cases, thereby highlighting the distinctions. In the second section, I begin a review of the empirical evidence with an assessment of how researchers typically measure monetary policy shifts. The following two sections examine the methods used for differentiating between the theories. Studies fall into two broad categories depending on whether they use aggregate or disaggregate data. The third section discusses the aggregate data, while the fourth section describes the use of disaggregate data. A conclusion follows.

MONETARY POLICY: THEORY

A General Framework

One way of posing the fundamental question associated with understanding the monetary transmission mechanism is to ask how seemingly trivial changes in the supply of an outside asset can create large shifts in the gross quantity of assets that are in zero net supply. How is it that small movements in the monetary base (or nonborrowed reserves) translate into large changes in demand deposits, loans, bonds and other securities, thereby affecting aggregate investment and output?

The various answers to this puzzle can be understood within the framework originally proposed by Brainard and Tobin (1963). Their paradigm emphasizes the effects of monetary policy on investor portfolios, and is easy to present using the insights from Fama's (1980) seminal paper on the relationship between **financial**lt; intermediation and central banks.

Fama's view of **financial**lt; intermediaries is the limit of the current type of **financial**lt; innovation, because it involves the virtual elimination of banks as depository institutions. The setup focuses on an investor's **portfolio**lt; problem in which an individual must choose which assets to hold given the level of real wealth. Labeling the **portfolio**lt; weight on asset i as w sub i , and total wealth as w, then the holding of asset, i--the asset demand--is just X sub i = w sub i W.

In general, the investor is dividing wealth among real assets--real estate, equity and bonds--and outside money. Each asset has stochastic return, (character omitted), with expectation (character omitted); and the vector of asset returns, (character omitted) has a covariance structure (character omitted). Given a utility function, as well as a process for consumption, it is possible to compute the utility maximizing **portfolio**lt; weights. These will depend on the mean and variance of the returns, (character omitted) and (character omitted), the moments of the consumption process, call these mu sub c , and a vector of taste parameters that I will label and assume to be constants. The utility maximizing asset demands can be expressed as (equation omitted).(2)

This representation makes clear that asset demands can change for two reasons. Changes in either the returns process (character omitted) or

macroeconomic quantities (mu sub c , W) will affect the (characters omitted).(3)

At the most abstract level, **financial**lt; intermediaries exist to carry out two functions. First, they execute instructions to change **portfolio**lt; weights. That is, following a change in one or all of the stochastic processes driving consumption, wealth or returns, the intermediary will adjust investors' portfolios so that they continue to maximize utility. In addition, if one investor wishes to transfer some wealth to another for some reason, the intermediary will effect the transaction.

What is monetary policy in this stylized setup? For policy to even exist, some government authority, such as a central bank, must be the monopoly supplier of a nominally denominated asset that is imperfectly substitutable with all other assets. I will call this asset "outside money." In the current environment, it is the monetary base. There is a substantial literature on how the demand for outside money arises endogenously in the context of the type of environment I have just described.(4) But in addition, as Fama emphasizes, there may be legal requirements that force agents to use this particular asset for certain transactions. Reserve requirements and the use of reserves for certain types of bank clearings are examples.

Within this stylized setup, a policy action is a change in the nominal supply of outside money. For such a change to have any effects at all, (1) the central bank controls the supply of an asset that is both in demand and for which there is no perfect substitute, and (2) prices must fail to adjust fully and instantaneously. Otherwise, a change in the nominal quantity of outside money cannot have any impact on the real interest rate, and will have no real effects. But, assuming that the policymaker can change the real return on the asset that is monopolistically supplied, investors' **portfolio** weights must adjust in response to a policy change.

The view of **financial** intermediaries that is implicit in this description serves to highlight the Brainard and Tobin (1963) insight that monetary policy can be understood by focusing solely on the endogenous response of investor portfolios. Understanding the transmission mechanism requires a characterization of how asset holdings change in response to policy actions.

Second, even though there need be no banks as we know them, there will surely be intermediaries that perform the service of making small business loans. The agency costs and monitoring problems associated with this type of debt will still exist, and specialists in evaluation will emerge. While they will have such loans as assets, they most likely will not have bank deposits as liabilities. Such entities will be brokers, and the loans will be bundled and securitized.

With this as background, it is now possible to sketch the two major views of the monetary transmission mechanism. There are a number of excellent surveys of these theories, including Bernanke (1993a), Gertler and Gilchrist (1993), Kashyap and Stein (1994a) and Hubbard (1995). As a result, I will be relatively brief in my descriptions.

IHE MONEY VIEW

The first theory, commony labeled the money view, is based on the notion that reductions in the quantity of outside money raise real rates of return.(5) This, in turn, reduces investment because fewer profitable projects are available at higher required rates of return-this is a movement along a fixed marginal efficiency of investment schedule. The less substitutable outside money is for other assets, the larger the interest rate changes.

There is no real need to discuss banks in this context. In fact, there is no reason to distinguish any of the "other" assets in investors' portfolios. In terms of the simple **portfolio** model, the money view implies that the shift in the (characters omitted) for all of the assets excluding outside money are equal.

An important implication of this traditional model of the transmission mechanism involves the incidence of the investment decline. Since there are no externalities or market imperfections, it is only the least socially productive projects that go unfunded. The capital stock is marginally lower. But, given that a decline is going to occur, the allocation of the decline across sectors is socially efficient.

This theory actually points to a measure of money that is rarely studied. Most empirical investigations of monetary policy transmission focus on M2, but the logic of the **portfolio** view suggests that the monetary base is more appropriate. It is also worth pointing out that investigators have found it extremely difficult to measure economically significant responses of either fixed or inventory investment to changes in interest rates that are plausibly the result of policy shifts. In fact, most of the evidence that is interpreted as supporting the money view is actually evidence that fails to support the lending view.

THE LENDING VIEW: BALANCE SHEET EFFECTS

The second theory of monetary transmission is the lending view.(6) It has two parts, one that does not require introduction of assets such as bank loans, and one that does. The first is sometimes referred to as the broad lending channel, or **financial** accelerator, and emphasizes the impact of policy changes on the balance sheets of borrowers. It bears substantial similarity to the mechanism operating in the money view, because it involves the impact of changes in the real interest rate on investment.

According to this view, there are credit market imperfections that make the calculation of the marginal efficiency of investment schedule more complex. Due to information asymmetries and moral hazard problems, as well as bankruptcy laws, the state of a firm's balance sheet has implications for its ability to obtain external **finance**. Policy-induced increases in interest rates (which are both real and nominal) can cause a deterioration in the firm's net worth, by both reducing expected future sales and increasing the real **value** of nominally denominated debt. With lower net worth, the firm is less creditworthy because it has an increased incentive to misrepresent the riskiness of potential projects. As a result, potential lenders will increase the risk premium they require when making a loan. The asymmetry of information makes internal **finance** of new investment projects cheaper than external **finance**.

The balance sheet effects imply that the shape of the marginal efficiency of investment cunre is itself a function of the debt-equity ratio in the economy and can be affected by monetary policy.(7) In terms of a simple textbook **analysis**, policy moves both the IS and the LM curves. For a given change in the rate of return on outside money (which may be the riskless rate), a lender is less willing to **finance** a given investment the more debt a potential borrower has. This points to two clear distinctions between the money and the lending views--the latter stresses both the distributional impact of monetary policy and explains how seemingly small changes in interest rates can have a large impact on investment (the **financial** accelerator).

Returning to the **portfolio** choice model, the presence of credit market imperfections means that policy affects the covariance structure of asset returns. As a result, the (characters omitted) will shift differentially in response to monetary tightening as the perceived riskiness of debt issued by firms with currently high debt-equity ratios will increase relative to that of others.(8)

IHE LENDING VIEW: LOANS FROM INTERMEDIARIES

The second mechanism articulated by proponents of the lending channel can be described by dividing the "other" assets in investors' portfolios into at least three categories: outside money, "loans" and all the others. Next, assume that there are firms for which loans are the only source of external funds--some firms cannot issue securities.(9) Depending on the solution to the **portfolio**amp; allocation problem, a policy

action may directly change both the interest rate and the quantity of loans. It is not necessary to have a specific institutional framework in mind to understand this. Instead, it occurs whenever loans and outside money are complements in investor portfolios; that is, whenever the **portfolio**amp; weight on loans is a negative function of the return on outside money for given means and covariances of other asset returns.(10)

The argument has two clear parts. First, there are borrowers who cannot **finance**amp; new projects except through loans, and second, policy changes have a direct effect on loan supply. Consequently, the most important impact of a policy innovation is cross-sectional, as it affects the quantity of loans to loan-dependent borrowers.

Most of the literature on the lending view focuses on the implications of this mechanism in a world in which banks are the only source of loans and whose liabilities are largely reservable deposits. In this case, a reduction in the quantity of reserves forces a reduction in the level of deposits, which must be matched by a fall in loans. The resulting change in the interest rate on outside money will depend on access to close bank deposit substitutes. But the contraction in bank balance sheets reduces the level of loans. Lower levels of bank loans will only have an impact on the real economy insofar as there are firms without an alternative source of investment funds.

As a theoretical matter, it is not necessary to focus narrowly on contemporary banks in trying to understand the different possible ways in which policy actions have real effects. As I have emphasized, bank responses to changes in the quantity of reserves are just one mechanism that can lead to a complementarity between outside money and loans. As pointed out by Romer and Romer (1990), to the extent that there exist ready substitutes in bank portfolios for reservable deposits such as CDs, this specific channel could be weak to nonexistent.(11) But it remains a real possibility that the optimal response of investors to a policy contraction would be to reduce the quantity of loans in their portfolios.

The **portfolio**amp; choice model also helps to make clear that the manner in which policy actions translate into loan changes need not be a result of loan rationing, although it may.(12) As Stiglitz and Weiss (1981) originally pointed out, a form of rationing may arise in equilibrium as a consequence of adverse selection. But the presence of a lending channel does not require that there be borrowers willing to take on debt at the current price who are not given loans. It arises when there are firms which do not have equivalent alternative sources of investment funds and loans are imperfect substitutes in investors' portfolios.

Obviously, the central bank can take explicit actions directed at controlling the quantity of loans. Again, lowering the level of loans will have a differential impact that depends on access to **financing**amp; substitutes. But the mechanism by which explicit credit controls influence the real economy is a different question.(13)

DISIINGUISHING IHE TWO VIEWS: GENERAL CONSIDERATIONS

Distinguishing between these two views is difficult because contractionary monetary policy actions have two consequences, regardless of the relative importance of the money and lending mechanisms. It both lowers current real wealth and changes the **portfolio**amp; weights.(14)

Assuming that there are real effects, contractionary actions will reduce future output and lower current real wealth, reducing the demand for all assets. In the context of standard discussions of the transmission mechanism, this is the reduction in investment demand that arises from a cyclical downturn.(15)

The second effect of policy is to change the mean and covariance of expected asset returns. This changes the (character omitted). In the simplest case in which there are two assets, outside money and everything else, the increase in the return on outside money will reduce the demand for everything else. This is a reduction in real investment.

The lending view implies that the change in **portfolio**amp;

weights is more complex and in an important way. There may be some combination of balance sheet and loan supply effects.

This immediately suggests that looking at aggregates for evidence of the right degree of imperfect substitutability or timing of changes may be very difficult. What seems promising is to focus on the other distinction between the two views--the lending view's assumption that some firms are dependent on loans for **financing**amp;.

In addition to differences stemming from the relative importance of shifts in loan demand and loan supply, the lending view also predicts cross-sectional differences arising from balance sheet considerations. These are also likely to be testable. In particular, it may be possible to observe whether, given the quality of potential investment projects, firms with higher net worth are more likely to obtain external funding. Again, the major implications are cross-sectional.

EMPIRICAL EVIDENCE: PRELIMINARIES

Before discussing any empirical examination of the monetary transmission mechanism, two questions must be addressed. First, do nominal shocks in fact have real effects? Unless monetary policy influences the real economy, it seems pointless to study the way in which policy changes work. Second, how can we measure monetary policy? In order to calculate the impact of monetary policy, we need a quantitative measure that can reliably be associated with policy changes.

Here I take up each of these issues. In the following section, I will weigh the evidence on the real effects of money. This is followed by a discussion of ways in which recent studies have attempted to identify monetary shocks.

THE REAL EFFECTS OF NOMINAL SHOCKS

Modern investigation of the impact of money on real economic activity began with Friedman and Schwartz (1963). In many ways, this is still the most powerful evidence in support of the claim that monetary policy plays an important role in aggregate fluctuations. Through an examination that spanned numerous monetary regimes, they argue that apparently exogenous monetary policy actions preceded output movements.

Recent researchers use more sophisticated statistical tools to study the correlations between money and income. This "money-income causality" literature is largely inconclusive, because it fails to establish convincingly either that money "caused" output or the reverse. In the end, the tests simply establish whether measures of money forecast output, not whether there is causation. Given that outside money--the monetary base--is less than 10 percent of the size of M2, it is not surprising that economists find the simultaneity problems inherent in the question too daunting and give up.

Two pieces of evidence seem reasonably persuasive in making the case that money matters. First, the Federal Reserve seems to be able to change the federal funds rate virtually without warning. (I am not arguing that this is necessarily a good idea, just that it is possible.) In the very short run, these nominal interest rate changes cannot be associated with changes in inflationary expectations, and so they must represent real interest rate movements. Such real interest rate changes almost surely have an impact on real resource allocations.(16)

The second piece of evidence comes from the examination of the neutrality of money in Cecchetti (1986, 1987). In those papers, I establish that output growth is significantly correlated with money growth at lags of up to 10 years! There are several possible interpretations of these findings, but they strongly suggest that monetary shocks have something to do with aggregate real fluctuations.

MEASURING INNOVATIONS TO MONETARY POLICY

It stands to reason that before one can study the monetary transmission mechanism, it is necessary to identify monetary shocks. A number of authors have argued convincingly that policy disturbances cannot be gauged by examining movements in the monetary aggregates. The reason is that the variance in the innovations to broad measures of money are a

combination of endogenous responses to real shocks (King and Plosser, 1984) and shifts in money demand (Bernanke and Blinder, 1992).

There have been two reactions to the fact that monetary aggregates provide little insight into policy actions. Both begin by looking at the functioning of the Federal Reserve and examining how policy is actually formulated. The first, due to Bernanke and Blinder (1992), note that the federal funds rate is the actual policy instrument that is used on a day-to-day basis. This suggests that innovations to the federal funds rate are likely to reflect, at least in part, policy disturbances. The main justification for their conclusion comes from examining the institutions of how monetary policy is carried out.

Romer and Romer (1989) suggest a second method. By reading the minutes of the Federal Open Market Committee (FOMC) meetings, they have constructed a series of dates on which they believe policy became contractionary.(17)

Innovations to the Federal Funds Rate

To understand the shortcomings of these two approaches, I will describe how each is used. In the first, researchers begin by specifying a vector autoregression. For the purposes of the example, I will use the formulation in Bernanke and Blinder's (1992) Section IV. They employ a six-variable specification with the total civilian unemployment rate, the log of the CPI, the federal funds rate, and the log of three bank balance sheet measures, all in real terms: deposits, securities and loans. The assumption is that the federal funds rate is a "policy" variable, and so it is unaffected by all other contemporaneous innovations.(18)

Following Bernanke and Blinder, I estimate the VAR with six lags using seasonally adjusted monthly data.(19) Figures 1 and 2 plot some interesting results from this VAR. (Figures 1 and 2 omitted.) The first figure shows the estimated residuals from the federal funds rate equation. The solid vertical lines are National Bureau of Economic Research (NBER) reference cycle peaks and troughs, while the dashed vertical lines are the Romer and Romer dates, intended to indicate the onset of contractionary monetary policy episodes.

This series looks extremely noisy and it is hard to see how it could represent policy changes. The 1979-82 period is the only one with large positive or negative values. Although it is surely the case that there are unanticipated policy changes both when the Federal Reserve acts and when it does not, one would expect small normal shocks with occasional spikes. If decisions are really this **random**, there is something fundamentally wrong with the policymaking apparatus. Furthermore, since the federal funds rate itself is the equilibrium price in the reserves market, given technicalities of the way that monetary policy is actually carried out, the market-determined level of the funds rate is not a policy instrument.(20)

The second figure shows the response of the log of the CPI to a positive one percentage point innovation in the federal funds rate. To understand how this is computed, begin by writing the vector autoregression as

A(L)y sub t = e sub t ,

where A(L) is a matrix of polynomials in the lag operator L (1quation omitted) is the vector of variables used in the estimation, and e sub t is mean zero independent (but potentially heteroskedastic) error. The first step is to estimate the reduced form version of equation 1 by assuming that no contemporaneous variables appear on the right-hand side of any equations (A(O) = I). This results in an estimate A(L) along with an estimated covariance matrix for the coefficient estimates--call this Omega. The impulse response functions are obtained by inverting the estimated lag polynomial B(L)= A sup -1 (L).(21)

But the point estimate of the impuise response function is not really enough to allow us to reach solid conclusions. It is also important to construct confidence **intervals** for the estimates. There are two ways to do this. The first involves the technique that has been called Monte Carlo Integration. This is a Bayesian procedure that involves

presuming that the distribution of the vector of errors in equation 1--the e sub t 's--is i.i.d. normal.(22) To avoid making such stringent assumptions, I choose to estimate confidence bands using an alternative technique grounded in classical statistics.

The delta method is the simple procedure that comes from noting that if the estimates of the coefficients in lag polynomials are asymptotically normally distributed, then any well-behaved function of these parameters will also be asymptotically normally distributed. Stacking all of the parameters in A(L) and calling the result Theta, then

(Equation omitted.)

It follows that any function of these parameters f(Theta)--for example, the impulse response function--will be asymptotically normally distributed,

(Equation omitted.)

where

(Equation omitted.)

which can be estimated numerically.

The result plotted in Figure 2 was first pointed out by Sims and is known as the "price purzle." Paradoxically, the VAR estimates imply that monetary policy contractions lead to price increases! As is clear from the estimated standard-error bands, this price rise is significantly positive for approximately the first year. After two years, however, it is not possible to reject the hypothesis that a funds rate increase has no effect on the price level.(23)

The standard conclusion is that the VAR is misspecified in some way. One strong possibility is that the funds rate is not exogenous in the way that is required for this identification to be valid, and so these innovations do not accurately reflect policy movements.(24)

My conclusions may be too harsh for the following reason. As Ben Bernanke pointed out in the conference, the estimated innovations are the sum of true policy innovation, policy responses to omitted variables, and more general specification errors in the VAR. As a result, one would expect them to be noisy. Furthermore, as pointed out by Adrian Pagan, since one is primarily interested in the impulse response functions--the impact of unanticipated policy on output, prices and the like--then it may be immaterial that the estimated policy innovations are noisy even if the true innovations are not.

The Romer and Romer Dates

The Romer and Romer dates have been both widely used and extensively criticized.(25) They suffer from both technical and substantive problems. First, they are discrete. Presumably, policy changes have both an intensity and a timing. Ignoring the size of policy changes must have an impact on results. Second, Romer and Romer choose to focus their inquiry only on policy contractions, because they feel that expansions were more ambiguous. Since most models predict symmetric responses to positive and negative monetary innovations, this strategy throws out information.

But the main issue is the exogeneity of the policy shifts. It is difficult to believe that the actions of the FOMC, as reported in the minutes of the meetings, are truly exogenous events. There have been two responses to this. First, Hoover and Perez (1991) provide a lengthy discussion of why Romer and Romer's methods are not compelling in identifying output fluctuations induced by exogenous monetary shocks.

Taking a slightly different approach, Shapiro (1994) examines whether the FOMC is responding to changes in economic conditions, and so there is some reaction function implicit in policy. He estimates a probit model for the Romer and Romer dates using measures of inflation and unemployment, both as deviations from a carefully constructed target level, as determinants. Figure 3 reproduces his estimates of the probability of a date, with the vertical lines representing the dates themselves. (Figure 3 omitted.) The unanticipated policy action is 1 minus the estimated probability. As is clear from the figure, several of the dates were largely anticipated, and there were some periods when policy shifts were thought to

be likely, and then did not occur. Overall, Shapiro's results suggest that the standard interpretation of the dummy variables as exogenous is incorrect to varying degrees over time.

There seems to be no way to measure monetary policy actions that does not raise serious objections. Given this, it might seem difficult to see how to proceed with the study of different theories of the transmission mechanism. But the literature proceeds in two directions. The first uses these measures directly in an attempt to gauge the influence of policy changes directly. The conclusions of these studies must be viewed with some degree of skepticism. The alternative approach is to note that investment declines account for the major share of output reductions during recessions. If one is able to show that the distribution of the contraction in investment is correlated with variables related to a firm's balance sheet and its access to bank loans, then this strongly suggests the existence of a lending channel.

USING AGGREGATE DATA

Numerous studies have used aggregate data in an attempt to distinguish the channels of monetary transmission. This literature can be divided into three caegories: The first looks at the relative forecasting ability of different quantity aggregates; the second studies differences in the timing of the response of aggregate quantities to presumed policy shocks; and the third examines the behavior of interest rates.

Before examining the work on quantities, I will discuss the use of interest rate data.(26) As is clear from the discussion in the first section, the lending view does allow for movements in market interest rates. Furthermore, these movements are in the same direction as those predicted by the money view, and their magnitude depends solely on the degree of substitutability between outside money and various other assets. Where the two views differ is in their predictions for movements in the interest rate on loans. But since there is currently no secondary market for these securities, it is impossible to determine the interest rate on these loans.(27) This implies that market interest rates are of virtually no use in this exercise. There is no sense in which the behavior of interest rates could serve to distinguish between the money or lending views.

I now turn to the work on quantities. In the following section, I examine tests involving the relative forecasting ability of measures of money and credit. This is followed by a discussion of papers that emphasize aggregate timing relationships.

Relative Forecasting Ability

A number of papers have examined the ability of different **financial**amp; aggregates to forecast output (or unemployment) fluctuations. Ramey (1993) is a recent example. The main methodology here is to ask whether measures of credit are informative about future output movements, once money has been taken into account. The problem with this is that credit is usually just a broader measure of money. To put it slightly differently, the balance sheet identity of the banking system implies that bank assets equal bank liabilities. As Bernanke (1993b) points out, monetary aggregates are a measure of bank liabilities, while credit aggregates are measures of bank assets. Since these are calculated slightly differently, they will not be identical. But it is these technical measurement differences that are likely to account for the differences in forecasting ability, not anything about the transmission mechanism.

More generally, the main finding is that credit lags output. Unfortunately, this tells us nothing about the transmission mechanism. The aggregate data do show that aggregate credit is counteucyclical, but it is easy to find explanations for this that are consistent with the lending view. For example, Kiyotaki and Moore (1993) present a model in which individuals must continue to service credit even after income falls, and so credit falls after income even though it is the fundamental source of fluctuations. In the end, it is difficult to see how aggregate timing relationships can tell us anything at all about the way in which monetary

policy affects real activity.(28)

Aggregate Timing Relationships

The second use of aggregate data has been to examine the response of various **financial**amp; quantities to policy innovations. Returning to Bernanke and Blinder (1992), they study whether bank loans and securities respond differently to federal funds rate innovations.(29) The standard methodology is to calculate the impulse responses for the two variables and note that they look different. Figure 4 reports the common finding, calculated using the six-variable Bernanke and Blinder VAR estimated over the 1959-90 sample. (Figure 4 omitted.) In response to a positive 1 percentage point innovation in the federal funds rate, the unemployment rate rises by nearly 0.1 percentage point after one-and-a-half years, while bank securities fall 0.07 percent and loans decline 0.02 percent. Securities fall both by a larger amount and more quickly than loans.

But point estimates of these impulse responses do not tell the entire story. In Figure 5, I plot the point estimate and two standard error bands for the difference between the impulse response for loans and securities. (Figure 5 omitted.) This allows an explicit test of whether these two assets are imperfectly substitutable in response to the shock. The differences are individually greater than zero in only a few months, and a joint test of the first 24 months of the impulse response, which is asymptotically distributed as a Chi-squared, has a p-**value**amp; of 0.70.

My conclusion is that reduced-form vector autoregressions are nearly incapable of providing convincing evidence of a differential impact of federal funds rate innovations on various parts of bank balance sheets. These results are based on the estimation of a large number of parameters with a relatively small amount of data--this VAR has 237 parameters and 354 data points--and so the estimates are fairly imprecise.(30)

But even if one were to find that the impulse responses differed significantly, this would only bear on the substitutability of the assets, and not directly on the validity of the lending view. Both the prices and quantities of perfect substitutes must have the same stochastic process, and so finding that this particular partial correlation is different would be evidence of imperfect substitutability. As Bernanke and Blinder (1992) make clear in discussing their findings, this is a necessary but not a sufficient condition for the lending view to hold. It is not possible, using reduced-form estimates based on aggregate data alone, to identify whether bank balance sheet contractions are caused by shifts in loan supply or loan demand. What is needed is a variable that is known to shift one curve but not the other.

Kashyap, Stein and Wilcox (1993) also provide evidence based on aggregate timing. They compare the response of bank loans to that of commercial paper issuance following policy innovations. They find that monetary policy contractions seem to decrease the mix of loans relative to commercial paper. Borrowers that can move away from direct bank **finance**amp; following a tightening appear to do so. Both Friedman and Kuttner (1993), and Oliner and Rudebusch (1993) take issue with these findings and show that changes in the mix are due to increases in the amount of commercial paper issuance during a recession, but that the quantity of bank loans does not change. In addition, Oliner and Rudebusch show that once firm size is taken into account, and trade credit is included in the debt of the small firms, the mix of **financing**amp; is left unaffected by policy changes.

It is worth making an additional point about the commercial paper market. First, Post (1992) documents that all commercial paper rated by a rating agency must have a backup source of liquidity, which is generally a bank line of credit or a standby letter of credit. This means that commercial paper is an indirect liability of banks, albeit one that is not on their balance sheet. Furthermore, Calomiris, Himmelberg and Wachtel (1994) suggest that increases in commercial paper issuance are accompanied

by an increase in trade credit. This means that a policy contraction may simply cause a re-shuffling of credit by forcing banks to move liabilities off of their balance sheet such that large firms issue commercial paper in order to provide trade credit to small firms that would have otherwise come from banks.

USING CROSS-SECTIONAL DATA

There is a large empirical literature using cross-sectional data that is relevant to understanding the channels of monetary policy. These studies fall into groups that separately address the two parts of the lending view. The first set of papers tries to gauge the importance of capital market imperfections on investment, and so is related to the balance sheet effects described in the first section. The second set, which is fairly small, examines time-series variation in cross-sectional data in an attempt to characterize the distributional effects of monetary policy directly. I will briefly describe each of these strategies.

Measuring Captial Market Imperfections

The literature on capital market imperfections is an outgrowth of the vast work done on the determinants of investment. The general finding in this literature is that internal **finance**lt; is less costly than external **finance**lt; for firms that have poor access to primary capital markets.

The empirical studies fall into two categories. The first examines reduced-form correlations, while the second looks directly at the relationship between the cost and expected return to a marginal investment project--they estimate structural Euler equations.

Reduced-Form Correlations

Fazzari, Hubbard and Petersen (1988) pioneered the technique of dividing firm-level data into groups using measures thought to correspond to the project monitoring costs created by information asymmetries, and then seeing if the correlation between investment and cash-flow measures varies across the groups. The finding in a wide range of studies is that investment is more sensitive to cash-flow variables for firms who have ready access to outside sources of funds.(31)

The main issue in interpreting these results is whether the characteristics of the firm used to split the sample are exogenous to **financing**lt; decisions. Measures of firm size, dividend policies, bond ratings and the like may be related to the quality of investment projects a firm has available, and so lender discrimination may not be a consequence of asymmetric information.

There are several examples in which researchers identify potentially constrained firms based on institutional characteristics, and so the endogeneity problems are mitigated. I will mention two. Hoshi, Kashyap and Scharfstein (1991) find that investment by Japanese firms that were members of a heiretsu, or industrial group, was not influenced by liquidity effects. Using data on individual hospitals, Calem and Rizzo (1994) find that investment depends more heavily on cash-flow variables for small, single-unit hospitals than for large, network-affiliated ones.

In the most convincing study of this type, Calomiris and Hubbard (1993) study the undistributed corporate profits tax in 1936 and 1937 to estimate the differences in **financing**lt; costs directly from firms' responses to the institution of a graduated surtax intended to force an increase in the dividend payout rate. Their results, holding investment opportunities fixed, are that investment spending is affected by the level of internal funds only for those firms with low levels of dividend payments and high marginal tax rates. Furthermore, these tended to be smaller and faster growing firms.

Structural Estimation

The neoclassical theory of investment allows one to derive the complex equilibrium relationship among the capital stock, rates of return, future marginal **value**lt; products and project costs that form the first-order conditions for a firm's problem. With the appropriate data, it is then possible to see whether these Euler equations hold. Hubbard and

Kashyap (1992) is an interesting use of this technique. Following the work of Zeldes (1989) on consumption, they examine whether the ability of agricultural firms to meet this first-order condition depends on the extent of their collaterizable net worth. They find that during periods when farmers have high net worth, and so have better access to external **financing_** xmlns:rates="urn:x-prefix:rates" xmlns:investment="urn:x-prefix:investment" xmlns:dbh=

Special Features: Graphs; Equations; References
Classification: 1120 (CN=Economic policy & planning); 1110 (CN=Economic conditions & forecasts); 8100 (CN=Financial services industry); 9190 (CN=United States); 9130 (CN=Experimental/Theoretical)
Descriptors: Federal Reserve monetary policy; Economic impact; Securities markets; Regulation of **financial** institutions; Studies
Geographic Names: US

36/9/17 (Item 17 from file: 268)
00258182
Strategic implications of the changing consumer marketplace
Nagel, Tracy L
Retail Banking Digest , v 15 , n 1 , p 4-23 , Jan/Feb 1995 **Document Type:** Newsletter Article **Article Type:** Feature **Journal Code:** BACB **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 08192

Abstract:

To understand demographic trends, bankers must develop a knowledge of the underlying social and cultural origins of individual market segments. Five key trends have emerged that impact the demographic reconfiguration of today's consumer marketplace. These trends reflect an increase of: 1. an aging population, 2. ethnicity and cultural diversity, 3. the number of single parent families, 4. consumers marrying and having families later in life, and 5. mobility and transience of the consumer marketplace.

Text:

Introduction

The composition of the consumer marketplace is changing dramatically--extending far beyond the phenomenon of the aging Baby Boomer generation. In general, today people are expected to live longer, grow up with a single parent, marry and have children later in life, and work for multiple employers over the course of their careers.

To understand demographic trends, bankers must develop a knowledge of the underlying social and cultural origins of individual market segments. Equally important are the socioeconomic, political, and technological conditions that most influenced consumers as **financial** behavior and habits were solidified. These behavioral drivers, coupled with

demographic shifts, create a need to rethink traditional approaches to **financial** services.

Marketers and managers must be able to develop an understanding of their markets from much more than just a traditional age and income perspective. Generational, cultural, and family origins have significant influences on when, how, and to what degree a consumer conducts business with a bank. These demographic phenomena will impact significantly the way banks develop products and services and create the innovative delivery mechanisms required to better serve the needs of the marketplace in the future.

Much has been published about the changing dynamics of the consumer marketplace. Many trends have been identified as a result of extensive research and **analysis** of U.S. Census data as well as consumer, social, and economic research by leading demographers, market researchers, and sociologists. This article draws on much of these earlier analyses.

The first two sections summarize the key findings from several of these studies. The third section focuses on the strategic and management implications these trends will have on the banking industry. Publications and authors are referenced throughout these sections. A "suggested reading" overview with full bibliographical information follows this article.

Demographic Trends

Five key trends have emerged that impact the demographic reconfiguration of today's marketplace. Specifically, these trends reflect an increase of--*an aging population;

* ethnicity and cultural diversity;
* the number of single parent families;
* consumers marrying and having families later in life; and
* mobility and transience of the consumer marketplace.

Aging Population. No one demographic trend receives more attention than this: People are living longer. Currently, people aged 65 and older comprise less than 13% of the population. By the year 2025, this segment is expected to grow to represent 20% of the population.(1)

Medical innovations, an increasing focus on health and fitness, and new technologies have made sustaining life into the mid-eighties a reasonable expectation. Demographers forecast growth in the 65-plus age bracket during the next ten years and expect a dramatic, significant shift by the year 2011, when the first Baby Boomers reach the age of 65.(2)

In 1990, approximately 9 million Americans aged 65 or older were living alone. These numbers are expected to grow to approximately 13 million households by the year 2010, with an estimated 75% of these households expected to be women.(3) Part-time care and assistance as well as carefully designed wellness plans make such independent living feasible.

Retirement center and community living have mass appeal to this market segment. The huge increase in the over-65 population will sustain significant growth in retirement villages, where aging consumers can live out their lives and enjoy the recreational activities and travel opportunities typically available.

Aging consumers will have an increased need to plan for retirement. The average Baby Boomer is expected to live well into his or her eighties, and many will live beyond that.(4) Comprehensive **financial** plans will be needed to stretch retirement-based investments and to support life-style maintenance and asset preservation. A large portion of these plans must include provisions for medical coverage and care, as well as the transition to a more controlled living environment.

The well publicized impact of this growth on the Social Security fund has resulted in lower consumer expectations. Concerns regarding the long-term sustainability of the Social Security system have forced many aging consumers to save more and use alternative sources, such as 401(k) plans, to supplement their income requirements. Independent investment plans are also growing in popularity among these demographic segments to compensate for these expected changes.

Aging consumers still consider banks their primary source for filling

deposit, credit, and investment needs. Today's retired consumers will continue to represent the core deposit base of most banks. They tend to use the bank as an investment source, preferring the bank's perceived image as a safe, conservative place to put their funds.

The branch is the primary delivery channel used to support the transaction and servicing needs of retired customers. They represent the majority of consumers who pass through a branch during daytime business hours. Banks should consider locating branches in retirement communities, with hours, amenities, and personnel tailored to consumer demand and personal service. The concept may provide opportunities to capture significant wallet share from this segment, allowing banks to allocate resources and qualified sales staffing an environment that is highly responsive to personal interaction.

The attitudes and behavior of the aging market will change considerably. Baby Boomers who are more comfortable with technology today will remain comfortable with technology as they age. **Financial** goals of security and risk aversion are likely to be consistent with today's retired; however, the method of delivery will change significantly.

Increasing Ethnicity and Cultural Diversity. The United States was founded, established, and built as a nation of immigrants. **Historically**, over time these immigrants tended to meld into a relatively homogeneous population. Today's "melting pot" of immigrants are more likely to maintain their ethnic and cultural distinctions, creating unique marketing and product development challenges. These challenges may prove difficult to overcome but are usually well worth the effort.

Population projections reflect the rapid growth and changes in America's ethnic profiles. According to the U.S. Census Bureau, the number of U.S. Hispanics has nearly doubled from 1980 and is estimated to reach 34 million by the year 2000. Projections are that the Hispanic population could surpass the number of African Americans in the next twenty to thirty years. The Census Bureau projects that, by the year 2050, Hispanics will represent more than one-fifth of the population(5) (see Exhibit 1). (Exhibit 1 omitted) Today, 14% of the population speaks a language other than English.(6)

Asians, Hispanics, and African Americans comprise an estimated $300 billion market.(7) Bankers must be prepared to approach ethnic markets in a different manner. For example, research indicates that, when compared to other segments of the same ethnic group, where ethnic consumers tend to live implies specific attitudinal and behavioral distinctions. This suggests a need for banks to avoid "mass market" marketing strategies. Instead, extensive research will be required to effectively target ethnic markets and to understand individual market and behavioral differences.

Local market knowledge is critical to understand the varying cultures and **financial** product, service, and delivery demands. For example, some ethnic groups, particularly the more affluent, view their banker as a close, trusted advisor. This behavior is common with new arrivals and the nonresident alien market, whose expectations are relative to their native country. In these instances, the relationship often extends beyond traditional account service and transactions. Many expect their banker to help them find doctors, lawyers, and insurance agents, and even to recommend an appropriate place to shop. In contrast, other ethnic segments are wary of banks as a provider of any product, regardless of the service. Banks must, therefore, evaluate carefully the potential profit contribution based on the time and cost required to meet and respond to the highly individualized service and relationship expectations of these consumer segments.

Communication and sales strategies for ethnic markets may also present challenges. Marketing strategies must go beyond a simple translation of an advertisement into a foreign language. Effective programs will reflect an understanding of how the promotion will be interpreted.

Strong, personal relationships will result from developing appropriate, focused communication programs. Employees in highly

concentrated ethnic markets should have bilingual and/or multilingual capabilities. In addition, bank management should be active participants in local markets' cultural events.

Increasing Number of Single-Parent Families. Since 1960, the percentage of single-parent households has more than tripled(8) (see Exhibit 2). (Exhibit 2 omitted) In 1988, an estimated 26% of all births were by unmarried women. Only 41% of all children born today are likely to reach maturity living with both biological parents.

Clearly, for such single-parent households, time and place convenience and affordable **financial** services are critical. Supermarket branching, extended hour telephone banking, and enhanced ATM functionality strategies may be effective in market areas with high concentrations of single-parent families. These consumers are likely to place less emphasis on a highly personal relationship with their bank or banker when selecting providers.

Access to information and advice will be critical to single-parent households in order to facilitate objective-based planning efforts. Of primary concern in developing **financial** plans is **financing** their children's education as well as funding any additional activities their children wish to be involved with (for example, sports, special schools, and so on). Banks must devise processes to deliver information to consumers who require assistance but may be unwilling or unable to pay.

Consumers Are Marrying and Starting Families Later in Life. Married couples represented 55% of all households in 1990, down from 60% in 1980.(9) Education and career goals are at the forefront of planning, dictating life-style and family structure with many younger segments. The median age for a women to marry is approximately 24 years old, almost 4 years older than the 20 years of age that was typical for a woman to marry in 1955. Men are also marrying later, with the median age of approximately 26 today, also up four years from 22 in 1955.(10)

Couples marrying later in life leads to starting families later. As a result many families are having fewer children. The average family size is 2.6 people, a decline from 3.3 people 25 years ago(11) (see Exhibit 3). (Exhibit 3 omitted)

Smaller family households can be attributed in part to the increasing number of single-parent families and older people living alone. Today, the average age of a first-time mother is about 30. At the age of 18, a child's parents are more likely to be approaching retirement. Families, therefore, are simultaneously **financing** college and planning for retirement, requiring significant amounts of information and advice to reach **financial** planning objectives .

Consumer Migration and Mobility. Consumers are moving more than ever before. Younger consumers are seeking quality job opportunities outside of the market where they grew up or went to college. Retired consumers find the warmer climates of Florida and Arizona an escape from the cold Northeast and Midwest winters.

Most people who move from state to state tend to have higher education levels and are somewhat more affluent. The move is typically a result of a career or advanced educational opportunity or retirement. While some families relocate in support of the female's career opportunities, household mobility is still largely in response to opportunities for the male household member.

Unmarried people find it easier to move and represent a large portion of the younger migration base. Migration patterns fluctuate, however, depending on the socioeconomic conditions of the marketplace.

Consumer migration will have an important impact on interstate branching. Before, moving meant changing banks and establishing new **financial** relationships. In the future, banks will be positioned to leverage opportunities with existing customers when they move, allowing them to take their **financial** relationships with them. Banks must develop the appropriate programs (for example, relocation packages) as an

important strategy to effectively serve and retain customer relationships.

Generational Trends

Demographic trends are demonstrative of a larger marketplace: a shift in dominating generations. Generational **analysis** is a study of the behaviors and attitudes of market segments based largely on time of birth and the influence of socioeconomic and political events. Attitudes and behaviors are strongly influenced by the social, economic, and political issues present during people's formative years.

Consumers are also affected by behavioral traits and characteristics: where, how, and what type of life-style an individual is born into, grows up in, and matures into will determine strongly his or her behavior.(12) These issues and **historical** events shape the attitudes and behaviors of each generation. These attitudes, in turn, have a significant impact on when, where, and to what degree each generation will interact with banks and other **financial** service providers.

Most industry experts agree that there are five "living" generations. While the names given to each generation vary across the experts, birth years and general life spans tend to be similar. The time frame of each generation is estimated at approximately 20-year **intervals** (see Exhibit 4). (Exhibit 4 omitted) William Strauss and Neil Howe, the authors of Generations and 13th Gen, are referenced throughout this section, as are several other professionals who have analyzed generational characteristics and behaviors.

The Last of the Traditionalists. The oldest of the five generations, the Traditionalists, are people born anywhere from the turn of the century to about 1925. Traditionalists were born during WWI and grew up during the days of Prohibition and the Great Depression. They witnessed the roaring 20s and fought in WWII. They are the last to grow up in the period of grand elegance, social graces, and class distinctions that determined the "status" of every family.

As a result, the Traditionalists tend to have high social and moral standards, tend to be very conservative in nature, and are most resistant to dramatic changes in society and the American culture. Traditionalists tend to believe that everything should move forward in an orderly fashion; they do not understand or tolerate chaos. They grew up during America's growth into an economic power--when America became a dominant producer of goods and services nationally and internationally. Younger Traditionalists were the first Boy Scouts and Girl Scouts in America. As adults, they were the first to buy the "blue chip" stocks of IBM, Xerox, and AT&T.

This generation, dubbed the "G.I.'s" by Strauss and Howe, experienced the greatest increase in educational attainment compared to any other before them. "The average length of in-school schooling rose from the ninth grade level to the twelfth, the share of 20-year-olds attending college tripled, and math and science aptitudes rose sharply."(13) The G.I. Bill also provided opportunities for education that were not available to earlier generations, making education affordable.

In addition, Traditionalists are thought to be the most affluent elders in our history; they are the only elders of the 20th century. They tend to have higher rates of home ownership, health insurance coverage, discretionary income, and net worth compared to others.(14)

Traditionalists, in total, hold a significant amount of wealth accumulated during the post-WWII era. They are the first generation to benefit from Social Security disbursements and significant pension plans. Housing values and equity provided additional benefits, especially during the inflationary 1980s.

As a result, Traditionalists are expected to pass on significant levels of wealth to surviving generations over the next two decades, leaving large inheritances to their children, grandchildren, and great grandchildren. This intergenerational wealth transfer is estimated to reach $10 trillion.(15) Assets currently held in trust funds and estates will be transferred, leaving the surviving generations with substantial decision-making responsibilities regarding investments. This inheritance

will serve as a retirement investment base for many, to compensate for years of heavy spending and accumulated credit debt.

Traditionalists are perhaps the most loyal to banks and trust companies of any generation. Their savings behavior is most influenced by the low levels of inflation and actual deflation during their younger years. They did not have to worry about the eroding **value** of their savings. Traditionalists are more concerned with the safety of their principal. Banks appeal to their general conservative nature and unwillingness to assume risk. They tend to live on the interest and dividend income they receive in addition to their pension and Social Security income and prefer not to use their asset base to support their life-style needs. To continue to meet the needs of this generation, banks will have to rely on the trusted, advisory relationship and provide investment management and transfer assistance.

Not surprisingly, Traditionalists are today's primary users of trust services, and **financial** and investment management services have a broad appeal to them. Basic deposit products are the Traditionalists' preferred liquid savings vehicle. They are the generation who consider the passbook savings account most valuable. They have no need for credit, due in part to the lack of credit availability during their younger, working years. Traditionalists have already made their major life purchases of home, furniture, appliances, and the like. Any mortgages owed by this generation are long paid off or carry very low outstanding balances.

The branch is the primary delivery channel used by the Traditionalist. This generation firmly believes that in-person service is the only proper manner to conduct business with a bank. They are easily frustrated with the telephone and mistrust the automated teller machine (ATM). Checks and cash are the main payment options used by Traditionalists.

The Depressionists. This generation was impacted by the highly charged political era of the Great Depression and World War II. Born somewhere between 1925 to about 1945, Depressionists are considered by some to be an "in between" generation; born in time to suffer through the hardships of the Great Depression and witness the rebirth of America under Franklin Roosevelt's New Deal, but too young to fight in the second World War. They were the youngest generation to witness the devastation that nuclear weapons brought and the birth of the Cold War. They were the fighters of the Korean War.

Those born in the earlier years of this generation have started to retire. Strauss and Howe believe the "silent generation" has "enjoyed a lifetime of steadily rising affluence, have suffered relatively few war casualties, and have the twentieth century's lowest rates for almost every social pathology of youth (crime, suicide, illegitimate births, and teen unemployment)."(16) Consistent with the turbulent political times of their youth, they represent virtually every prominent figure in the civil rights movement, including Martin Luther King.(17)

The G.I. Bill has allowed this generation to excel and surpass others in the area of educational achievement. Depressionists have a greater tendency to have worked for one company during their professional careers. Corporate loyalty and commitment guaranteed almost any professional a steady climb up the ladder. The smaller size of this generation relative to others allowed for more opportunities to compete and advance in the job market. Their loyalty has been rewarded with a large pension and retirement after 30 years of service.

Depressionists' conservative **financial** behavior is influenced most by the economic strains and the political events of their childhood and young adulthood. They are similar to the Traditionalists in their unwillingness to assume risk that may jeopardize the true **value** of their asset base. Depressionists have a general mistrust of **financial** services providers; some will, as a result, keep portions of money hidden in mattresses, books, and so forth, in the event of economic downturns.

Although Depressionists are expected to pass on decreased levels of

wealth compared to Traditionalists, they will be contributors to the huge intergenerational wealth transfer. However, longer life expectancies and high medical costs will impact the total **value** of the wealth they accumulated during their working and primary asset generation years.

Banks continue to be the preferred **financial** services provider compared to other alternatives. Depressionists place a primary emphasis on asset preservation to maintain the real **value** of their money through retirement. A strong demand for information and advice will emerge to support investment management decisions due in part to increased life expectancies. Allocation and preparation for high medical costs must be considered when developing **financial** plans. Depressionists' low-risk approach to investing make fixed-income security instruments, such as corporate bonds and government securities, attractive product alternatives.

Traditional, basic deposit services have mass appeal to this generation. Interest checking and savings products are valuable to them; they are similar to Traditionalists in that they also like passbook savings accounts. Along with the Traditionalists, they will continue to supply banks with their core deposit base.

Depressionists were the first generation to have broad availability of and subsequent demand for credit. Non-mortgage installment credit developed in the 1930s as a way to increase car sales. Mortgages facilitated home purchases, and later home equity lending made vacation homes and other recreational activities as well as **financial** planning possible. Credit cards and charge accounts were attractive although somewhat risky to this generation. Access to credit in retirement years will continue to be important to the Depressionist generation. Usage of and propensity to carry balances will decline as credit becomes more of a convenience, not a necessity or a way to liquify illiquid assets. Mortgages will be paid off; smaller outstanding balances may be held in retirement housing. Credit cards and home equity lines of credit will be a way to access funds.

Like the Traditionalists, Depressionists are largely branch-dependent and have high service demands. Their conservative nature and general mistrust of banks make them question almost everything. Depressionists also tend to be extremely price-sensitive and demonstrate a general unwillingness to pay for service. Packaged accounts with qualifying balances to offset service charges are attractive.

Depressionists born in the 40s may be somewhat more accepting of ATM and telephone technology. However, the vast majority of this generation is uncomfortable with, and therefore unwilling to use, alternative electronic delivery channels.

The Baby Boomers. Perhaps the most talked about and overanalyzed generation was born roughly between the years 1945 and 1964. Named for the post-World War II population boom, this one generation represents the largest portion of all living Americans. According to William Dunn in The Baby Bust, "Births reached 3.4 million in 1946 and kept climbing, breaking through 4 million by 1954, and peaking at 4.3 million in 1957. For 11 straight years, births topped 4 million."(18) An estimated 77 million people were born during this era, and they are the focus of almost every cultural, social, and political issue discussed today.

The generation growing up in the post-war economic, manufacturing, and production boom became known as the generation of great expectations. As children, they were perhaps the most nurtured of any generation; only about 2% attended institutional child care as preschoolers.(19) On the other hand, as children they became painfully aware of the concept of divorce.

Boomers were the first generation to grow up with television. Many of their social values were formed and influenced by the mass media. The television served as a messenger of the social unrest of the 1960s; the place to which they fled upon learning of John F. Kennedy's assassination and stayed for three days. They grew up knowing of the realities of nuclear attack, the Cold War, and the Berlin Wall. Boomers were the first generation (and last) to participate in air raid drills.

Boomers became adults during the tumultuous 1960s; they were the flower children and hippies and makers of the sexual revolution. The Boomer generation divided on the issue of the Vietnam war: those who went to fight and those who stayed home to question. According to Generations, "Only one Boomer man in sixteen ever saw combat. Among all the rest, two-thirds attributed their avoidance to some deliberate dodge."(20)

Boomers survived the decade of the 60s and transitional 70s and emerged in the 1980s as the "yuppies." They have been characterized as on the fast track, materialistic, selfish, and seeking instant gratification. Many secured high-salaried professional jobs with large corporations and have come to represent the majority of middle management. They were the power earners and spenders. However, corporate restructuring, layoffs, and cutbacks have made the 1990s rather unpleasant for the Boomers. Boomers are being replaced by the younger technology-oriented cheaper labor of the generation that followed.

The first Boomer is expected to reach the age of 65 around 2010. Demographers and forecasters are waiting to see how the Boomers will redefine retirement. Many are expected to exit the work force and establish their own small business ventures. Still others will take what little they have saved and the inheritances from Traditionalists and Depressionists and live a life of leisure.

Boomers have only recently realized the strong need to save, and those who are approaching retirement years have had to change their spending habits. They have a strong need for information and advice to facilitate their asset generation demands. Boomers differ from Traditionalists and Depressionists in that they are more willing to take calculated risks in return for a higher yield when investing. They are more comfortable with making decisions on their own.

Boomers are the heaviest users of credit, both secured and unsecured, to **finance** their lifestyle needs. They are more willing and likely to carry balances than preceding generations. Credit will continue to play an active role in the Boomer's **financial** life-style into and throughout retirement. However, the role of credit will shift. Revolving balances are expected to decline, auto leasing options will become more attractive, and home equity credit will be accessed to fund recreational activities as well as their children's college educations.

Boomers distinguish themselves from other generations in their demand for technology and innovation. They were the first generation to use ATMs actively and now place frequent calls to voice response units and use screen phones and computers to manage their **finances**. They are among the least branch-dependent consumers. The Boomers' life-style requires the convenience of time and place that traditional brick and mortar facilities can't offer.

Boomers' attitudes reflect a decreasing dependence on traditional commercial banks. They are the first generation to actively use nonbanks for **financial** services. They shop actively for the affordably priced product that meets their expectations in terms of access and management. Boomers' affinity for technology makes product specialists, such as investment firms and mortgage and **finance** companies, attractive alternatives.

Boomers respond to a "system of doing business" that meets their life-style goals and needs. This is the generation that is redefining banking as a distribution rather than a product business. As a result, Boomers are more likely to fragment relationships, leaving banks with a relatively small to insignificant portion of their wallet.

Generation X. The generation immediately following the Baby Boomers has come of age in the eyes of analysts and forecasters. This generally forgotten and misunderstood generation of consumers was born from 1965 through 1982. Also known as the baby busters, lost generation, twentysomethings, 13th generation, and the generation after, Generation X appears to be the name that has been associated most often with their group.

Xers are difficult to describe due to their highly individualistic behavior. They are the most unique of all living generations in terms of their cultural mix, family structure, and socioeconomic environment. According to Generations, "...Thirteeners were America's true children of the 60s and especially the 1970s. An awakening era that seemed euphoric to young adults was, to them, a nightmare of self-immersed parents, disintegrating homes, schools with conflicting missions, confused leaders, a culture shifting from G to R ratings, new public health dangers, and a Me Decade economy that tipped toward the organized old and away from the voiceless young."(21)

Generation Xers are the first children parents took pills not to have. Divorcing parents significantly impacted this generation, and the number of single-parent families has risen dramatically. They are the first generation to be significantly impacted by step parenting and extended families. According to Generations, "...the number of latchkey children under the age of 14 left alone after school roughly doubled."(22) They are the first generation to become aware of and exposed to AIDS.

Alone, without parental supervision or nurturing most of the time, Xers turned to the television as their babysitter, educator, and mentor. They grew up with programs like Sesame Street. Television served as their window to the world and exposed them to the social unrest of the times. Here they witnessed Watergate, Nixon's resignation, the Iran hostage crisis, and Contra hearings. Xers do not remember the assassination of John F. Kennedy like the preceding Boomers. However, almost anyone born in this generation remembers where they were when the space Shuttle Challenger exploded upon lift-off--most Xers were watching it on television.

Electronics and technology occupied this generation, marked by the entrance of Atari, the Game Boy, and the video arcade. Computers were introduced and, ultimately, were in widespread use by Xers in schools as they grew up.

Education levels for Xers are generally lower compared to other generations. According to Generations, "The 13th (generation) is on its way to becoming the first generation...to be less college educated than its next elders. College completion rates, seven years after high school graduation, fell from 58 percent of the Baby Boomer class of 1972 to 37 percent of the 13er class of 1980."(23)

The opportunities for advanced education, while more readily available, are less affordable. Many have been forced to work full-time to **finance** their education and attend college on a part-time basis. Of those who have graduated, the majority have exited college with substantial student loans and heavy **financial** burdens.

The labor market has not been extremely favorable to the Xer. The Baby Boomers continue to dominate the shrinking middle management segment of the corporate organizational structure. Economic conditions and heavy overhead burdens have presented a difficult challenge for Xers to secure and retain employment commensurate with their education and training. Many Xers have returned to their parents' home as a result of being forced into lower paying jobs to gain work-related experience. Generation Xers are the first generation that is not expected to surpass the standard of living of its preceding generation.

Generation Xers' needs for **financial** services are relatively basic. Many do not have the high balance requirements to qualify for more sophisticated, yield-oriented deposit instruments. They tend to have low balance, high transaction activity, basic checking accounts. They tend to use ATMs to make small withdrawals.

Xers have very few opportunities to save money. Generation Xers were not taught the concept of savings by parents or schools. As a result, many believe that money can only be saved in lump sums, and they have a need to understand the **value** of gradual savings. If Xers are saving at all, their 401(k) plan is the primary investment source.

Growing up in the shadow of the Baby Boomers, they learned and mastered the concept of spending on credit. Their strongest credit product

demands are credit cards and auto loans. Xers, in many instances, cannot afford or qualify for lower rates and fee structures, forcing them to pay interest rates on revolving balances. Xers are the first generation to have credit card access during their college years. Many card issuers relaxed their qualification criteria to capture highly profitable **finance** charges from revolving balances. As a result, Xers have exited college with high revolving credit balances in addition to student loan burdens. Home ownership is not an option for most Xers at this stage in their lives. Many expect this generation to purchase homes later in life compared to other generations.

Most Generation Xers do not recall the last time they visited a bank branch. To this generation, the ATM is the branch. They are the generation most responsive to electronic, alternative delivery channels. They are uncomfortable with personal interaction, preferring the telephone, ATM, computer, and other technologies as their communication, service, and transaction vehicles.

The Millenials. America's youth of today will be tomorrow's drivers in the fast lane of the Information Superhighway. This generation is expected to include all children born through the end of this century and into the third millennium. Born in 1982, the class of 2000 brings the high hopes and expectations of a new America.

Millennials mark a return to a more caring, nurturing childhood environment. The impact of the social unrest of the 60s and 70s and the economic conditions of the 80s have caused many adults to evaluate the consequences of their actions and return to foster more human, caring values. Millennials are growing up in an era of refocus; society is changing the way it thinks about and reacts to violence, gangs, drugs, and sex. These are the children of "just say no" campaigns and AIDS. Unlike others before them, Millennials are a "wanted" generation and are expected to grow up in families who planned for them and were ready to take responsibility for their upbringing.

Millennials will be most influenced by the access to information through emerging technologies. Advancements in CD-ROM technology and the affordability of enhanced computer systems make the computer the new television of the 90s. Controlled educational games, books, and other activities create an environment that promotes the advancement of their young minds. Millennials' comfort with technology is evident; in fact, many are teaching the elder Boomers how to use the systems.

Millennials are also expected to save more compared to previous generations. According to Generations, "The per capita savings rate for children age 4 to 12 held steady at about 15 percent from 1968 to 1984--but by 1989 it had risen sharply, to over 30 percent. In the early 1980s, kids that age saved only one-third as much as they spent on candy, soft drinks and snacks. Today, with the encouragement of their parents, they save more than they spend on convenience food."(24)

From a banking perspective, products will become less important, and the method of access will become the leading distinction and basis for relationship development with this generation. Electronic delivery providing 24-hour access will be required to meet the needs of these consumers. Millennials will not use the branch and will be less impressed with personal service.

Expected to save more compared to their Boomer and Xer predecessors, Millennials will be the benefactors of every mistake made by their elder generations' spending habits. While the Millennials' demand for credit will most likely be consistent with earlier generations, the manner in which they use it is likely to be different. They will most likely carry lower revolving balances on credit cards or pay off outstandings monthly.

Millennials will also be more comfortable with debit and smart card options. In-person interaction will be replaced by electronic capabilities. More than any earlier generation, Millennials trust machines, in some cases, over people. Cash and check payments will be replaced with electronic transfer and debit transactions. Branches will not be thought

necessary by this generation.

Implications for Strategy and Implementation

These five generational trends will be the basis of a period of transition for the **financial** services industry. Over the next five to ten years, the consumer marketplace will experience significant, dramatic changes. The marketplace will be impacted sharply by both demographic phenomena and shifts in generational attitudes (see Exhibit 5). (Exhibit 5 omitted)

This transition period must not become an excuse to delay the major changes in business practices, products, customer service and sales processes, and delivery systems required to win in this new consumer world. It will take substantial time and effort to put these changes in place and to develop the skills required to manage and retain customer relationships in tomorrow's marketplace. Delay will allow bank and nonbank competitors to build skills and delivery systems that cannot be overcome five years from now. **Financial** providers need to start planning today to meet the primary strategic imperatives that tomorrow's plans must consider:

* increased population diversification;
* shifts in consumer segmentation;
* dual requirements for life-style maintenance;
* distribution preference migration;
* information hunger; and
* competition against nonbanks.

Increased Population Diversification. The consumer marketplace is more ethnically, culturally, morally, and socially diverse than ever before. Key demographic trends will impact how and when consumers conduct business with their **financial** providers (see Exhibit 6). Traditional approaches to banking strategy will not suffice in a consumer marketplace that is essentially continuously reinventing itself (see Exhibit 7). (Exhibit 7 omitted)

Shifts in Segmentation. The uncontrollable change in the population's dynamics, coupled with the emergence of new technology and information-oriented consumers, will require a need to reevaluate standard segmentation approaches. Traditional segmentation schemes will become increasingly difficult to manage as the market differentiates itself into subgroups based on a multitude of demographic, behavioral, and generational characteristics.

Simply put, today's consumer behavior is not what it will be tomorrow. Today's mass marketing strategies must mature into "market of one" segmentation approaches to appeal to the highly diverse and unique demands of the market (see Exhibit 8). (Exhibit 8 omitted) Traditional, one-size-fits-all strategies will not work in tomorrow's marketplace. Successful strategies will integrate a process to design highly flexible information **value**-added products and create innovative delivery options to serve the wide-ranging needs of customers. Product preferences will be replaced by distribution demands.

Customers will want the ability to design their own relationships with their bank, selecting the products and services they need from a broad product menu of options. A high degree of management focus and discipline will be required to allocate and prioritize resources, capital, and investments to the most critical, high-impact areas. Successful **financial** providers will empower customers to design and manage their banking relationships, using emerging electronic, nonsite-based delivery options (see Exhibit 9). (Exhibit 9 omitted)

Dual Requirements for Lifestyle Maintenance. **Historically**, as consumers aged, they tended to save more and spend less. This "life cycle" savings behavior has served as the foundation for many **financial** providers' strategic planning and development efforts. However, the new aging generations will create a demand that reflects their desire to continue spending to support their life-style as well as save for the future. This trend represents a new duality to the traditional "life cycle" behavior.

Comfort with and dependence on credit is likely to continue into retirement. However, consumers will use credit differently in retirement, most likely to even out their spending habits. For example, the use of leasing will serve as a credit device to keep life-style intact.

Growing concerns about the longevity of Social Security and employer-based pension funds has forced consumers to focus on savings. Gradual savings plans and a sharp focus on investment objectives and life-style requirements will become critical components in developing a sustainable investment base to carry consumers through retirement. But astute borrowing will be a part of retirement living. Preparing for the consumers' dual needs and developing programs to incent and encourage both borrowing and savings behavior will be important in the future **financial** environment.

Distribution Preference Migration. Tomorrow's **financial** services environment will be distribution, not product, dominated. Competitive advantage and differentiation will be driven by accessibility options and product excellence. Superior accessibility will count for more than product features or complexity when attracting and retaining customers in this new environment. Key components of tomorrow's distribution system will include:

* convenient, responsive delivery channels designed to meet consumers' wide-ranging needs;

* tailored systems of doing business that allow customers to self-select and define accessibility;

* organizational integration of business units where necessary to create an accessible environment for customers; and

* elimination of internal conflict where it creates barriers for the customer to enter or move around the distribution system.

The overall goal is to create a distribution system that is easy to enter, easy to move around in, and difficult to exit.

Obviously, this reflects what is already happening: The role of the bank branch is changing dramatically. Its influence as a delivery channel will decline correspondingly with demographic and generational shifts (see Exhibit 10). (Exhibit 10 omitted)

Bank management will need to evaluate carefully the most appropriate role of the branch in tomorrow's distribution system. For the most part, individual market areas will dictate their needs. For example, supermarket branching and limited transaction facilities may be appropriate in markets with high concentrations of younger, convenience-oriented consumers. Full-service branches may continue to be the most effective alternative in more mature markets.

The emergence of electronic technologies will bring channel alternatives such as screen phones and computer modems. Enhanced ATM functionality and telephone options will provide customers with new alternatives to manage their relationships. Whatever the combination, banks must consider that the successful distribution system of tomorrow will integrate various delivery channels and create a mix in which customers will be able to define their own service environment.

Information Hunger. Consumers now more than ever have a substantial need for information. And while they are deluged with information, they are ill-equipped to interpret it. They seek assistance in sorting it out and explaining how it relates to their planning needs.

Banks, to a large degree, have not been able to satisfy the consumer's appetite for advice. Although information is the critical component in decisions regarding such things as retirement planning, **financing** college educations, and planning for first-time home purchases, raw, unrefined data are confusing.

Successful **financial** providers will have a seat at the consumer's planning table. This will require developing the advisory skills necessary to facilitate and execute sound investment decisions. Information should be available through a variety of distribution options, again allowing consumers to self-select the degree to which they require these

services. Consumers have demonstrated a growing comfort with making their own **financial** decisions. Competitive advantage will be achieved according to the **value** information adds to the decisionmaking process and how consumers can access it.

Competing Against Nonbanks. The future marketplace may not distinguish between banks and nonbanks as new **financial** companies emerge. These companies will have banking and nonbanking origins that will influence their ability to deliver products and services to the new marketplace. Today, however, nonbanks have become the preferred provider to many consumers as a result of well-executed distribution strategies. Some, even wary, consumers have come to rely on nonbank providers to meet their demands for easily accessible, high-quality products at a price they are willing to pay.

Nonbanks have excelled in capturing market and wallet share from banks largely as a result of their ability to deliver products and services according to the customer's (not the institution's) preferences. Allowing customers to define the depth of their relationships and accessibility preferences will begin to close the widening gap on nonbank usage. Continued, traditional banking approaches will not stop the momentum nonbanks have built and leveraged to create a strong market position.

Information-Driven Strategies. Information will be the power tool in developing and implementing strategies in tomorrow's markets. Extensive analytical and educational efforts will be required to understand the dynamics and distinctions of marketing areas. New strategies will require integrating market, customer, and environmental information to understand and prepare for transitions in the marketplace.

Database marketing applications must be developed or enhanced to create innovative information systems that will allow banks to track, measure, and predict consumer behavior. Information warehousing and online access will empower employees who interact personally with the customer. Neural networks, predictive modeling, and other system enhancements can facilitate sales and retention efforts by providing information on the next product to target or the potential for customers to terminate their relationships.

Effective, objective-based planning efforts will result from an ability to integrate customer information with local market knowledge. Efforts will be enhanced by banks that can clearly manage the similarities and distinctions of customers relative to the market. Segmentation schemes that allow bankers to clearly understand the demographic composition and behavioral dynamics of the marketplace will empower local market management, providing bankers with the information they need to target and attract individual segments and individual consumers.

The Formula for Success

Planning for the transitions in the consumer marketplace must begin today. However, implementation does not need to take place overnight, and sudden action may, in fact, do more harm than good. Careful planning and transitioning processes will result in a strong and healthy position in tomorrow's market. Banks need to develop the process and a time line in order to manage the continuous reinvention of the consumer marketplace. The critical challenge is to manage a new, mixed marketplace. The winning strategy for managing in tomorrow' marketplace will--*meet the uncontrollable changes in the demographic composition of today's populations and the underlying generational shifts that will create a new marketplace by providing products and delivery options that fit the new life-styles of aging "borrower-savers";

* acknowledge information and distribution as the critical competitive advantage;

* install changes in business practices that result in information-based decisions;

* create a distribution mix that reflects customer demands for service, convenience, and technology;

* transition from broad-based segmentation approaches to market of

one; and

* rely on market knowledge and information systems as the foundation for developing actionable and attainable strategies.

Well-constructed strategic plans will empower **financial** services providers to transition effectively into tomorrow's marketplace. Developing a unique approach to meeting the dynamic needs of tomorrow's consumers today will increase the opportunities for creating a sustainable competitive position tomorrow.

However, planning needs to start today. The winning **financial** providers in tomorrow's market will be strategic innovators, not adopters, that are prepared for the challenges of a continuously changing consumer marketplace.

Exhibit 6

Implications for **Financial** Services of Demographic Trends

Demographic Trends Reflect an Increase of:--Implications on **Financial** Services

An aging population--Increased demand for information and investment advisory services to maintain life-style requirements.

Ethnic and cultural diversity--Targeted products to meet unique, culturally driven, **financial** demands.

Single-parent families--Focus on service delivery considering timed and place convenience demands.

Consumers marrying and having families later in life--Information and advisory service to meet multi-objective savings plans.

Mobility and transience of the consumer marketplace--Develop standardized products and flexible distribution systems that allow customers to easily move and transition relationships.

Notes

1 The Road to 2015, p. 128.

2 Ibid., p. 128.

3 American Demographics, August 1994, p. 40.

4 American Demographics, September 1994, p. 27.

5 The Road to 2015, p. 125-126.

6 Bank Marketing, November 1993, p. 65.

7 Ibid., p. 65.

8 The Index of Leading Cultural Indicators, p. 50.

9 Ibid., p. 56.

10 The Master Trend, p. 103.

11 American Demographics, August 1994, p. 59.

12 The Future Consumer, Frank Feather, p. 66.

13 Generations, William Strauss and Neil Howe, p. 267.

14 Ibid., p. 268.

15 American Demographics, September 1994, p. 26.

16 Generations, William Strauss and Neil Howe, p. 283.

17 Ibid., p. 284-285.

18 The Baby Bust, A Generation Comes of Age, William Dunn, p. 8.

19 Generations, p.305.

20 Ibid., p. 306.

21 Ibid., p. 321.

22 Ibid., p. 325.

23 Ibid., p. 326.

24 Ibid., p. 342.

Suggested Reading

Several authors and publications have been referenced throughout this article. However, a multitude of information has been published regarding these and other demographic and generational phenomena. The following highlights some of the key publications related to these trends.

Books:

Bennett, William . The Index of Leading Cultural Indicators: Facts and Figures on the State of American Society. New York, NY: Simon & Schuster, 1994.

Dunn, William. The Baby Bust: A Generation Comes of Age. Ithaca, New

York: American Demographics, 1993.
Feather, Frank. The Future Consumer. Toronto, Canada: Warwick Publishing Inc., 1994.
Howe, Neil and Strauss, William. 13th Gen: Abort, Retry, Ignore, Fail? New York, NY: Vintage Books, A Division of **Random** House, Inc., 1993.
Petersen, John L., The Road to 2015: Profiles of the Future. Corte Madera, CA: Waite Group Press, 1994.
Roberts, Sam. Who We Are: A Portrait of America Based on the Latest U.S. Census. New York, NY: Time Books, A Division of **Random** House, Inc., 1993.
Rushkoff, Douglas. The GenX Reader. New York, NY: Ballantine Books, 1994.
Russell, Cheryl. The Master Trend: How the Baby Boom Generation Is Remaking America. New York, NY: Plenum Press, 1993.
Strauss, William and Howe, Neil. Generations: The History of America's Future 1584 to 2069. New York, NY: William Morrow and Company, Inc., 1991.
Wolfe, David B. Marketing to Boomers and Beyond: Strategies for Reaching America's Wealthiest Market. New York, NY: McGraw-Hill, Inc., 1993.
In addition, publications such as American Demographics focus their analytical efforts on the state of the marketplace today and into the future. Anyone who is interested in learning more should consider adding any or all of these to their collection.

Tracy L. Nagel, Senior Consultant with Furash & Company, Washington, D.C., has a broad background in the areas of market research and information management including market segmentation, strategic planning, competitive monitoring, product development, database and direct marketing applications, MCIF development and enhancements, distribution systems **analysis**, and customer retention.

Prior to joining Furash & Company, Ms. Nagel was Marketing Information Administrator at Society Corporation, where she managed the collection and reporting of all market, competitive, and secondary information. In this capacity, she was responsible for development of the applications and models to support ongoing strategic efforts related to sales and segment management, customer profitability, competitive intelligence, direct marketing, and retention initiatives. Additionally, she worked extensively on the Society-Ameritrust merger, supporting database development for direct mail communications, implementation of a customer retention database measurement system, as well as managing the research department integration process. Earlier assignments with Society Corporation included market research analyst and management associate positions.

Ms. Nagel received her MBA and BA in business administration with a concentration in marketing from Baldwin-Wallace College, Berea, Ohio.

Furash & Company is a Service Member of the American Bankers Association. The firm can be reached at 2001 L Street, N.W., Washington, D.C. 20036. Phone: (202) 331-8400. Fax: (202) 331-8399.

Special Features: Charts; Graphs; References
Classification: 9190 (CN=United States); 8100 (CN=Financial services industry); 7100 (CN=Market research)
Descriptors: Demographics; Lifestyles; Bank marketing
Geographic Names: US

36/9/18 (Item 18 from file: 268)
00257107
Alternative measures of real estate performance: Exploring the Russell-NCREIF data base
Fisher, Jeffrey D
Real Estate Finance , v 11 , n 3 , p 79-87 , Fall 1994 **Document Type:** Journal Article **ISSN:** 0748-318X
Journal Code: RFN **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 04403

Abstract:

The Russell-NCREIF (National Association of Real Estate Investment Fiduciaries) Index (RNI) is by far the most widely cited and used measure of the investment performance of institutionally held commercial property in the US. The RNI is constructed each quarter from a database that contains a wealth of basic information about the nature and operating performance of the properties NCREIF members hold. Using this underlying database to construct several new and different measures that may be relevant to gauging real estate investment performance over time is discussed. These new measures are not necessarily better than the traditional RNI reported each quarter by NCREIF, but they do offer some new insights and cast a somewhat different light on the recent history of the commercial property market. Details are provided.

Text:

The Russell-NCREIF Index (RNI) is by far the most widely cited and used measure of the investment performance of institutionally held commercial property in the United States. Indeed, this index is often referred to as the industry benchmark. Reported publicly each quarter, the RNI is based on the most recent appraised values and net operating income of virtually all the properties held for tax-exempt institutions by the major investment advisory ms in the country (members of the National Association of Real Estate Investment Fiduciaries (NCREIF)). The index currently includes 1,568 properties with aggregate appraised **value** over $20 billion, held by forty-seven advisory firms. What is less widely appreciated is that the RNI is constructed each quarter from a data base that contains a wealth of basic information about he nature and operating performance of the properties NCREIF members hold.

The purpose of this article is to use this underlying data base to construct several new and different measures that may be relevant to gauging real estate investment performance over time. These new measures are not necessarily better than the traditional RNI reported each quarter by NCREIF, but they do offer some new insights and cast a somewhat different light on the recent history of the commercial property market.(1) Specifically, we look at the following indexes:

1. An equal-weighted version of the RNI, in which the return to each property counts equally in determining the index return each quarter, rather than being weighted by the value of the property as in the official RNI.

2. An index of the percentage change in the total net operating income (NOI) of the properties in the index, rather than the percentage change in appraised value as in the official RNI.

3. An index of the current net operating cash flow yield of the properties in the index, a measure of the income return component defined slightly differently than that reported in the official index.

4. An index of the percentage change in the net operating cash flow (similar to #2), only net of capital improvement expenditures.

As we shall see, each of these ways of defining a real estate investment performance index based on the Russell-NCREIF data base complements the traditional published index.

REVIEW OF THE RUSSELL-NCREIF PROPERTY INDEX FORMULA

To understand the measures to be developed in this article, and how they differ from the published Russell-NCREIF Index, we begin with a brief review of how the official RNI is calculated. The following calculations are done for each property in the Russell-NCREIF data base for each quarter that the index is calculated:

Income Return =

(Equation (1) omitted)

Appreciation Return =

(Equation (2) omitted)

(3) Total Return = Income Return

+ Appreciation Return

where NOI refers to the net operating income of the property during the quarter; End Val refers to the appraised value reported by the advisor (or transaction price if the property was sold) as of the end of the quarter; Beg Val refers to the appraised value (of purchase price if the property was newly acquired) as of the beginning of the quarter; Cap Imp refers to capital improvement expenditures recorded being spent on the property during the quarter; and Part Sale refers to any amount received for partial sale of the property during the quarter.(2)

After the income, capital, and total returns are calculated for each property according to the equations, the returns are weighted by value. In effect, each property's return is multiplied by its appraised value for the quarter (actually, by the "denominator" described in the equation). The resulting product is then summed across all properties in the index data base at the beginning of the period, and the total is divided by the aggregate appraised value (or "denominators") of all the properties. This results in a value-weighted index in which larger properties (by value) carry a proportionately greater influence on the index return than do smaller properties.

The general rationale behind value-weighted indexes is that such weighting causes the index to measure the performance of a "market basket" reflecting the wealth of a complete population or portfolio. In this case the population or portfolio whose wealth is being measured is that of the commercial property holdings of the NCREIF membership tax-exempt clients. The index thus replicates the performance of all investment advisors put together.

EQUAL-WEIGHTED INDEX

Although the value-weighting used to calculate the RNI return is valid and interesting from the market basket perspective, from a statistical perspective **value**-weighting implies that the NCREIF properties are being viewed as a complete population rather than as a representative sample of some larger population of properties. This suggests that an alternative view, and a corresponding alternative way of defining the index return, might be of interest. If we view the properties in the Russell-NCREIF data base as a representative sample drawn from a larger population consisting of all institutional-grade commercial properties in the country (not just those currently held for tax-exempt clients by NCREIF members), then an equal-weighted index might be more appropriate than a **value**-weighted one. In an equal-weighted index each property's return would count equally in the calculation of the quarterly index return. The individual property returns would simply be summed across all properties and divided by the total number of properties in the index during that period. Thus, compared with the **value**-weighted

official RNI, the equal-weighted index is more strongly influenced by the relatively smaller properties.

Exhibit 1 shows the quarterly **historical** total returns to both the traditional **value**-weighted RNI and the corresponding equal-weighted index. (Exhibit 1 omitted) The actual quarterly return statistics are presented in the appendix. Although the two definitions of index returns give broadly similar results, the equal-weighted index (the broken line) has notably less short-run volatility than the **value**-weighted index (the solid line). Several explanations may lie behind this difference. From a purely statistical perspective, if all the properties are members of a single population and the quality of the appraisals is equal between lower-valued and higher-valued properties, then the equal-weighted index makes more efficient use of the data and so suffers less from **random** "noise" caused by idiosyncratic returns in a few large properties.

The two indexes also yield different results during periods of extreme shifts in property **value**. The equal-weighted returns ended to be lower than the RNI during he period of high returns up through the mid-1980s, but higher than the RNI during the slide in commercial property values in the late 1980s and early 1990s. This suggests that smaller properties have had a more stable long-term **value** history, less subject to the large swings in **value** that apparently have affected the larger properties more strongly. The governing factor behind this difference may not be property size per se, but rather location, with higher-valued properties tending to be located in central business districts (CBDs), where property **value** swings have been greater. Alternatively, the differences observed between the two indexes in Exhibit 1 could simply be because of the more frequent and careful appraisal of larger properties, resulting in less "smoothing" and lagging in appraisal-based returns.(3)

NOI CHANGE INDEX (PERCENT CHANGE IN NOI)

In both the traditional **value**-weighted RNI and the equal-weighted version, the appreciation return each quarter is based on the change in the appraised **value** between the beginning and end of the quarter. Thus, both of these indexes are based primarily on the asset market (i.e., the proper market) as opposed to the space market (i.e., the rental market). An alternative way to examine the change in investment performance, based on the space market, would be to consider the percentage change in the net operating income (NOI) between the beginning and end of the quarter.

An NOI change index, or percentage change in NOI, has several interesting features as a measure of investment performance. The NOI index would be of particular interest to long-term investors who are not contemplating selling their properties, but rather intend to hold them purely for the operating income they can provide. For such an investor, the NOI index will measure the change in productivity of the asset class. Another advantage is that a performance measure based purely on NOI would not be subject to appraisal smoothing (although NOI may be subject to other types or sources of smoothing). Asset markets are clearly of great importance to investors, but many academics and practitioners suspect that the property market sometimes misvalues the operating fundamentals of real estate assets, periodically overvaluing or undervaluing properties in a boom-bust cycle.(4)

Using an NOI index as a measure of investment performance does pose some disadvantages, at least in theory. The lack of a property market component reduces the relevance of the measure for ascertaining opportunity cost and making buy/sell investment decisions. NOI lacks the forward-looking element present in asset market values. Furthermore, even as a measure of the space market, NOI suffers from the fact that it reflects revenues from existing (vintage) leases as well as new leases. Thus, the NOI index is not a current rental market index. In addition, from a statistical perspective, any measure of flow occurring over an entire

interval of time will be subject to a type of temporal aggregation smoothing when differences in this measure are taken across time (as compared to "stock" measures, which in principle can be evaluated as of a single instant in time). An index of percentage change in NOI might be criticized in theory as a smoothed, rearview mirror perspective of investment performance. Nevertheless, this measure seems to offer sufficient potential interest to deserve consideration, if only to gain additional insight on the real estate asset class.

The formula for the NOI change index during period "t" is as follows:

(Equation (4) omitted)

where the summation is over the i = 1,..., N individual properties in the index. Contrasting this formula with both Equations (1) and (2) reveals the conceptual difference with the traditional RNI returns.(5) In the absence of direct information on NOI being published by NCREIF, it is useful to note that approximating the above formula is possible using information that is published in the Russell-NCREIF The procedure to do this is as follows:(6)

(Equation (5) omitted)

Whether one uses the exact formula, Equation (4), or the approximation based on public information, Equation (5), the period of time "t" can represent a quarter, a year, or any other calendar period. The quarterly NOI change index appears to be excessively volatile, perhaps because of the timing of income recorded over the year. Thus, for general purposes, evaluating the NOI index on an annual frequency--one that changes yearly--may be preferable.

It is informative to compare the NOI index with the appreciation return component of the Russell-NCREIF. This is done in Exhibit 2, which displays the annual frequency NOI change index together with the annual RNI appreciation return. (Exhibit 2 omitted) (Recall that the RNI appreciation return is the **value**-weighted change in property **value** .) Observe that the NOI index and appreciation return track each other very closely from 1979 through 1989, suggesting a relatively constant cap rate, or income return component. After 1990 the appreciation return is lower than the NOI change index, suggesting a repricing of real estate. Whether this repricing in the asset market is temporary or permanent remains to be seen.

Another feature of note in Exhibit 2 is that the NOI change index appears to be nearly as volatile as the RNI appreciation return component. The latter is, in turn, much more volatile than the RNI income return component. Some researchers believe that the extremely low volatility of the RNI income return component indicates that cash flow, or at least NOI, of commercial properties is very stable. The NOI change index reveals that this is not really so. Two factors account for the low volatility of the RNI income return component. Primarily, when measured over short time **intervals** (such as a quarter or a year), the NOI level is at least an order of magnitude smaller than the property **value** level. As the income return component is essentially a ratio of the NOI level divided by the property **value** level, this difference in magnitude causes large relative shifts in NOI to appear minimized in the income return. For example, equal percentage changes in NOI and in property **value** level will result in a manyfold larger change in the appreciation return component than in the income return component. A secondary reason why the income return lacks volatility is that the NOI and the property **value** both tend to change persistently across time in the same direction, as seen in Exhibit 2. This causes changes in NOI to be somewhat offset in the ratio that defines the income return.

The amount of volatility in the annual NOI change index is perhaps a bit surprising when one considers that NOI reflects the dampening or smoothing effect of vintage leases. In principle, one would expect NOI volatility to be much smaller than appreciation return volatility unless rents are highly mean-reverting and predictable.(7) Although not shown here, the NOI change index is quite volatile even for property types, such

as office buildings, where leases are predominantly long-term and fixed-rent. This suggests that real estate may be fundamentally volatile, i.e., that the observed volatility in property values is not merely revaluation in the capital market resulting from mood swings or preference shifts among investors. It is worth noting, however, that in spite of its rearview mirror nature, the NOI change index indicated an upturn in 1993 (positive gains in NOI level) while the RNI appreciation return still implied that the property market was headed downhill. The NOI change index also did not fall as far as the RNI appreciation return index in the early 1990s. Both of these trends seem to suggest an imminent turnaround in the property market.

CASH FLOW YIELD INDEX

The third index measure we examine concerns the income return component of the total return. As shown previously in the numerators of Equations (1) and (2), the RNI defines the income and appreciation return components such that the income return is based on he NOI without taking out capital improvement expenditures, while the appreciation return does not include the **value** of capital improvements made during the quarter (i.e., it subtracts this **value** out from the end-of-period property **value**). The result is that, from a cash flow perspective, the income return component overstates the income return, because capital improvement expenditures are cash overflows not available for distribution to investors. Similarly, the appreciation return understates the growth in actual property **value** over time because capital improvement expenditures increase the **value** of the property. These effects exactly offset each other in the total return, leaving the total return measure robust to this concern.

The argument behind the existing RNI definition of the income and appreciation components is based more on accrual accounting principles than cash flow. Nevertheless, to the extent that investors care about cash flow, examining a more purely cash flow-based measure of the income return component seems to provide some merit. This is particularly true as securitization of commercial property equity becomes more widespread, and analysts inevitably are tempted to draw parallels between the RNI income return measure and the cash flow-based measures of current yield prevalent in the stock market. Thus, the new index measure considered here is a redefined measure of income return that subtracts out the capital improvement expenditures from the NOI to yield a measure that is more similar to a current distributed cash flow yield indicator.

It is important to realize that capital improvements reported by members of NCREIF represent several different types of expenditures. A prominent type of expenditure that is often recorded as a capital improvement is what is referred to as a tenant improvement, or TI for short. Although TI may only improve the **value** of the building for a sine specific tenant, they are often not expensed out on a current basis (and so are not netted out of NOI). Another important type of capital improvement expenditure is general building renovation and rehabilitation, which improves the marketability and productivity of the building for all potential tenants for a long period of time to come. TIs on the one hand am part of the leasing deal and may be viewed as substitutes for lower rents. Therefore, subtracting such an expenditure from NOI to arrive at a cash flow number makes sense. Building renovations, on the other hand, are more like new investments, especially if they are major enough to involve additional infusion of capital from sources external to the building's own internally generated cash flow. In such cases, subtracting renovation expenditures from NOI in the year they are made would distort interpretation of the resulting cash flow as a measure of earnings from previous investment in the property. Unfortunately, differentiating between these two types of capital expenditures is not possible in the NCREIF data base.

To see what a cash flow yield would look like if we only subtracted capital expenditures other than major renovation, we can assume that

capital expenditures in excess of some arbitrary cutoff (here we shall use 20% of the appraised **value** of the property) are for major renovation. Thus, at the individual property level, capital expenditures in excess of 20% of property **value** are counted as reducing the appreciation return as in the traditional RNI formulation. Any other capital expenditures are subtracted from NOI to calculate the cash flow yield index. (The total return does not change.)

Exhibit 3 shows the difference between the RNI income return and our cash flow yield index for all the properties in the NCREIF data base. (Exhibit 3 omitted) The exhibit shows the annual yield measure, accumulated over four quarters. The cash flow yield index is naturally lower than the RNI income return. The cash flow yield index is also more volatile. The gap between the two indexes widened during the 1980s and was greatest in 1990. The average difference between the two return indexes is about 2 points per year, or some 200 basis points. This suggests that routine capital improvement expenditures including TIs amount to about 2% of property **value** per year, a sizable quantity. Thus, accounting for capital expenditures greatly changes the picture when comparing current yield measures, say, between property and REITs.(8)

CASH FLOW CHANGE INDEX (PERCENT CHANGE IN CASH FLOW)

The preceding two new index measures have examined, respectively, the percentage change in NOI as a purely fundamentals-based measure of performance, free of the asset markets, and the cash flow yield net of capital improvements, as an improved measure of yield based more on cash flow than accrual measures. The idea in this fourth, and last, new index measure is to combine these two approaches to calculate the percentage change in the cash flow net of capital improvements. The formula is depicted below, where CF represents the NOI minus all capital improvement expenditures for properties, and years in which those expenditures totaled less than 20% of the previous appraised **value**gt;.

(Equation (6) omitted)

Exhibit 4 compares the cash flow change index with the NOI change index discussed previously. (Exhibit 4 omitted) Clearly the cash flow index displays more volatility. This suggests hat the capital expenditures recorded here were probably not discretionary in the sense that they could not be easily deferred from one year to the next. If they had been, one would expect the property managers to use such spending flexibility to smooth out the property's cash flow and distributable income over time. Lack of temporal flexibility in capital expenditures, and the resulting increase in the volatility of the properties' net operating cash flow, is consistent with the hypothesis that most of the capital improvement expenditures recorded here are TIs or other such concessions necessary to compete in the normal rental market.

A legitimate question is whether the cash flow change index provides more information than the NOI change index, or just more noise. In this regard, note that the change in cash flow was negative in 1984, while the change in NOI was still positive. In retrospect, 1984 would indeed have been a good time to sell commercial property! Perhaps the dip in cash flow (due to a surge in TIs?) was indicating a change in the rental market fundamentals. Also, the change in cash flow was negative in 1990, while the change in NOI was positive. The cash flow measure may have given a more realistic picture during that recession year, when TIs were no doubt widely substituted for lower rents. In general,, the cash flow change index (and cash flow yield index) captures the use of TIs as a concession but the NOI change index (or RNI income return) does not. Thus, although the cash flow indexes should be interpreted with caution, they appear to provide useful additional information and therefore represent a nice complement to the purely NOI-based measures.

CONCLUSION

Clearly the Russell-NCREIF data base offers many opportunities for analyzing data besides the measures constructed and publicly reported in the traditional RNI. Looking a different kinds of measures and indexes

helps us unravel trends in the real estate market and better understand the past investment performance of the asset class. In particular, the four types of measures and indexes described in this article all complement the traditional Russell-NCREIF Index. I am hopeful that future research using the Russell-NCREIF data base will be able to explore each of the measures and indexes introduced in this article, and perhaps others as well, in more depth and detail. Possible research tasks include disaggregating the **analysis**gt; by property type and geographic area. Stay "tuned" to this journal for further reports on this ongoing research.

ENDNOTES

Portions of this article are based on material developed by Jeffrey D. Fisher for an NCREIF-sponsored seminar entitled "Real Estate **Portfolio**gt; **Analysis**gt; and Strategy and **Analysis**gt;." The opinions expressed in this article are solely those of the author.

1 By way of background, several articles have appeared recently in this journal discussing various aspects of the Russell-NCREIF Index. Brunette [1994] discusses the nature of the index and reports on metropolitan-level returns calculated from the underlying data base. Giliberto [1994] explains how the return formula used in the RNI was developed. Webb [1994] compares the transaction prices of properties sold from the data base to their most recent appraised values.

2 As explained by Giliberto [1994], in the Spring issue of this journal, the fractional parts of the capital improvements, partial sales, and NOI in the denominator on the right-hand side of the return definitions are included as a result of a mathematical approximation to the internal rate of return (IRR) over the quarter. This approximation assumes that the property is bought at the beginning of the quarter for the amount Big Val, sold at the end of the quarter for the amount End Val, and held during the quarter with NOI arriving in equal installments at the end of each month, and capital improvement expenditures and partial sales receipts occurring at the midpoint of the quarter. While this definition of the denominator in the return formula is somewhat unique among investment performance indexes, it is worth noting that the added fractional terms make little empirical difference in the aggregate index That is, the RNI return would be very much the same if the denominator were simply defined as the beginning **value**gt; of the property.

3 I am indebted to David Geltner, Mike Miles, and the participants in the NCREIF Seminar for suggesting the idea of the equal-weighted index and several of the hypotheses described here, and for explaining the differences between the equal-weighted and **value**gt;-weighted indexes. As with the other newly defined indexes in this article, the findings presented here represent only a first step in the **analysis**gt; of such alternative measures. I am hopeful that further research will shed more light on these and other hypotheses.

4 Studies such as Barkham and Geltner [1995] and Fisher et a. [1994] have found evidence strongly supporting the notion that private commercial property markets have lacked informational efficiency over the past decades in the U.S. It remains to be seen whether the increased securitization of commercial property equity via REITs and other vehicles will improve the informational efficiency performance of the private markets in the future.

5 Note, in particular, hat while both the income return of the RNI and the NOI index formula, Equation (4), are based on the NOI, the former views the NOI level as a fraction of the previous property **value**gt; . In contrast, the NOI index examines the change in NOI as a fraction of the previous NOI level.

6 The approximation would be exact except for the additional terms in the denominator of the RNI, discussed previously. Nevertheless, the approximation is very close to the exact **value**gt; in the Russell-NCREIF data base.

7 See Geltner [1990]. In addition, as we see shortly, even the NOI is smoothed compared to the net operating cash flow, after subtracting out

capital improvements, a measure of net cash that is arguably more relevant for the long-term investor.

8 In recent years REIT yields in the neighborhood of 6% have been compared with the RNI income return, suggesting much higher **valuation**gt; of property in REITs than in the private property markets an apparent arbitrage. But when one compares the yields on the basis of the cash flow yield index net of capital improvement expenditures, the gap narrows almost to nothing, even in 1993; in earlier years, REIT yields were clearly higher than private market yields.

REFERENCES

Barkham, R., and D. Geltner. "Price Discovery in American and British Property Markets." AREUEA Journal, Forthcoming, 1995.

Brunette, David. "Returns from Metropolitan Areas: Will 1994 Be a Rebound Year?" Real Estate **Finance**gt;, 11, 1 (1994).

Fisher J., D. Geltner, and R.B. Webb. "**Value**gt; Indices of Commercial Real Estate: A Comparison of Index Construction Methods." Journal of Real Estate **Finance**gt; and Economics, 9, 2 (Forthcoming 1994).

Geltner, D. "Return Risk and Cash Flow Risk with Long-Term Riskless Leases in Commercial Real Estate." AREUEA Journal, 18, 4 (1990).

Giliberto, S. Michael. "The Inside Story on Rates of Return." Real Estate **Finance**gt;, 11, 1 (1994).

Webb, R.B. "On the Reliability of Commercial Appraisals: An **Analysis**gt; of properties Sold m the Russell-NCREIF Index (1978-1992)." Real Estate **Finance**gt;, 11, 1 (1994).

Jeffrey D. Fisher is director of the Center for Real Estate Studies at the Indiana University School of Business in Bloomington, Indiana.

Special Features: Graphs; Appendix; Equations; References
Classification: 9190 (CN=United States); 8360 (CN=Real estate); 5200 (CN=Communications & information management); 3400 (CN=Investment analysis)
Descriptors: Real estate **financing**; Performance evaluation; Indexes; Return on investment; Commercial real estate; Data bases
Geographic Names: US

36/9/19 (Item 19 from file: 268)
00241866
Fair-**value** accounting for banks' investment securities: What do bank share prices tell us?
Barth, Mary E
Bank Accounting & Finance , v 7 , n 4 , p 13-23 , Summer 1994 **Document Type:** Journal Article **ISSN:** 0894-3958 **Journal Code:** ABAF **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 05423

Abstract:

Investment securities fair-**value** estimates are relevant to investors and sufficiently reliable to be associated with bank equity values. The relation between net assets before investment securities and bank share prices is significantly positive. More important, fair values of investment securities and unrealized investment securities

appreciation/depreciation also have a significant relation to bank share prices. Fair-**value** securities gains and losses appear less relevant and reliable than other earnings components, calling into question their appropriateness as a component of earnings. Fair-**value** investment securities not only have explanatory power incremental to **historical** costs in explaining bank share prices, but also **historical** cost had no explanatory power incremental to fair values.

Text:

As banks move to implement Statement of **Financial** Accounting Standards (SFAS) No. 115, "Accounting for Certain Investments in Debt and Equity Securities," some experts think the next target for mark-to-market accounting will be deposits. But should bank assets and liabilities be recognized in bank **financial** statements at fair **value**? Even if there are implementation and other issues relating to comprehensive fair-**value** accounting for banks, should at least investment securities be recognized at fair **value**? These fundamental questions underlie the fair-**value** accounting debate, particularly as it relates to banks.

This article summarizes the findings of a recent research study investigating air-**value**lt; accounting or banks' investment securities, used for calculating both assets and earnings. Although the study does not answer completely the questions posed above, it sheds some light on them. The study is based on data from approximately 150 large banks from 1971 to 1990. It finds that the estimates of investment securities fair values, disclosed in banks' annual reports, are reflected in bank share prices similarly to other bank assets. This suggests that investment securities fair-**value**lt; estimates are relevant to investors and sufficiently reliable to be associated with bank equity values. Moreover, the fair-**value**lt; estimates are more closely related to bank share prices than the **historical**_-cost xmlns:users="urn:x-prefix:users" xmlns:rea

Special Features: References; Graphs
Descriptors: Financial institutions; Securities; Stock prices; Fair **value**; Studies; **Historical** cost accounting; Regression **analysis**
Geographic Names: US

36/9/20 (Item 20 from file: 268)
00243928
Can deposit insurance increase the risk of bank failure? Some **historical** evidence
Wheelock, David C; Wilson, Paul W
Federal Reserve Bank of St. Louis Review , v 76 , n 3 , p 57-71 , May/Jun 1994
CODEN: FRBRDV **Document Type:** Journal Article **ISSN:** 0014-9187 **Journal Code:** FSL **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:

Word Count: 08071

Abstract:

The impact of deposit insurance is studied by drawing on **historical** evidence from a voluntary insurance system that operated in Kansas between 1909 and 1929. Because insurance was optional in this system, comparison of insured and uninsured banks facing otherwise similar regulations and economic conditions is possible. It is found that insured banks held less capital and reserves than uninsured banks, and that banks with low capital and reserves, or a heavy reliance on borrowed funds, were more likely to fail. In short, risky banks were more likely to fail, and members of the state deposit insurance system tended to be riskier than nonmembers.

Text:

LOAN LOSSES ASSOCIATED WITH declines in energy and agricultural prices, and the collapse of commercial real estate markets were the proximate cause of the high number of bank and savings and loan (S&L) failures of the past 12 years. Many researchers also blame government policies, however, such as restrictions on branch banking and limitations on the services that banks and S&Ls may offer. Such restrictions hamper diversification, thus leaving depository institutions particularly vulnerable to downturns in the regions which they serve.

Deposit insurance has probably been the most criticized government policy related to bank and S&L failures. Many economists believe that deposit insurance encourages banks and S&Ls to take excessive risks, thereby increasing their chance of failing.(1)

This article investigates empirically the connection between deposit insurance and bank failure. day, virtually all banks are insured by the Federal Deposit Insurance Corporation (FDIC) and, consequently, isolating the effects of insurance from other regulations and exogenous economic conditions that affect bank performance is difficult. We study the effects of deposit insurance by drawing on **historical** evidence from a voluntary insurance regime that operated in Kansas between 1909 and 1929. Because membership in the Kansas deposit insurance system was optional, we are able to compare insured and uninsured banks facing otherwise similar regulations and economic conditions in a way not possible with modern data. We estimate a model of bank failure to test for the impact of insurance on the likelihood of failure.(2) We find that insured banks were less well capitalized and, in some years, less liquid than uninsured banks, and that capitalization and liquidity were important determinants of failure.

The next section discusses how deposit insurance might increase the likelihood of bank failure. Next, we describe the Kansas deposit insurance system and the effects of a collapse of commodity prices in 1920 on commercial banks. The final sections develop the econometric methodology used to model failure, specify the model, and present results and conclusions.

DEPOSIT INSURANCE AND BANK FAILURE

Federal deposit insurance was enacted in response to the bank failures of the Great Depression. Thousands of banks failed from 1930 to 1933, wiping out the funds of depositors, producing a collapse of the money supply, increasing the costs of intermediation and interfering with the clearing of payments.(3) Although large banks and many economists vigorously opposed deposit insurance, and President Franklin Roosevelt was reluctant to accept it, Congress deemed deposit insurance necessary to

protect small, unsophisticated depositors from losses due to bank failures, and the payments system from a wholesale banking collapse like that of 1930-33.(4)

Until the 1980s, deposit insurance was generally hailed for eliminating the possibility of widespread bank failures.(5) Merton (1977) and Kareken and Wallace (1978) showed, however, that when insurance premiums are unrelated to the expected cost of failure to the insurance system, banks have an incentive to take greater risks than they otherwise would. Because depositors are protected in the event of bank failure (to the limit of insurance coverage), they have little or no incentive to monitor their banks' activities or to demand risk premiums on deposit interest rates. Deposit insurance thus raises the expected return to banks from investing in risky loans and investments and encourages them to substitute debt, in the form of insured deposits, for equity. Consequently, unless regulations inhibit risk-taking, the presence of deposit insurance could lead to more bank and S&L failures than there would otherwise be.

Many economists blame deposit insurance, coupled with inadequate regulation and supervision, political interference and a failure by regulators to promptly close insolvent institutions, for the high number of S&L failures and bankruptcy of the Federal Savings and Loan Insurance Corporation during the 1980s.(6) The banking industry's problems were, by comparison, less notorious. Banks faced higher capital requirements and were more stringently supervised than S&Ls, which lessened banks' incentive and ability to take excessive risks. But deregulation of deposit interest rates, initiated by the Depository Institutions Deregulation and Monetary Control Act (DIDMCA) of 1980, the gradual removal of barriers to branch banking, more liberal chartering polices and increased competition from foreign banks and from nonbank **financial** institutions, all worked to lessen charter values and increase the incentive for banks (as well as S&Ls) to take greater risks.' Moreover, in 1980, deposit insurance coverage was increased from $40,000 per account to $100,000 for both banks and S&Ls, while the failure resolution policy known as "too-big-to-fail" effectively extended insurance to all deposits at the largest banks, thereby enhancing their incentive to take risks.(8)

As is all too often the case, the bank and thrift debacle of the 1980s stemmed in part from the failure of policy makers to heed lessons from the past. Flood (1992) argues that when deposit insurance legislation was being considered in 1933, policy makers understood the temptation that insurance gives bankers to take excessive risks. Accordingly, coverage was limited to $2,000 per account and regulations were imposed to constrain risk-taking. Deposit interest rate ceilings prevented weak institutions from growing rapidly by bidding up interest rates, and regulators gave bankers added incentive to act conservatively by limiting the issuance of new bank charters. Many of the sources of increased competition for banks and S&Ls that had emerged by 1980, such as money market mutual funds and the commercial paper market, were the product of technological changes that could not be foreseen in 1933.(9) But deregulation of bank and S&L deposits and the expansion of deposit insurance coverage at a time when the industry was facing increased competition contradicted the regulatory principle that underlay deposit insurance legislation in 1933.

The insights that policy makers had in 1933 about deposit insurance came partly from prior state experiences with deposit insurance. Six states had experimented with insurance in the pre-Civil War era, as did eight others between 1908 and 1930.(10) None of the 20th-century systems was able to fully protect depositors of failed banks from loss, and each closed before the onset of the Great Depression. The commodity-price collapse of 1920-21 triggered a wave of bank failures throughout the Midwest and the South, including seven of the eight states with deposit insurance. Although loan losses associated with the decline of state incomes was the proximate cause of bank failures, insured banks generally suffered higher failure rates than uninsured banks facing similar exogenous conditions.(11) Contemporaries believed that deposit insurance had contributed to the high

number of failures because it protected incompetent and dishonest bankers from market discipline.(12) In the following sections, we investigate empirically how deposit insurance might have contributed to the failure of banks operating under the deposit insurance system of Kansas during the 1920s. We study this case because just three of the eight state insurance systems had optional membership for state-chartered banks and, hence, permit comparison of insured and uninsured banks facing otherwise similar conditions. Of these, only the Kansas system lasted for many years with a large number of banks electing to join the system and a significant number remaining uninsured.

DEPOSIT INSURANCE IN KANSAS

The Kansas deposit insurance system began operation in June 1909 and officially closed in 1929. Kansas was the second state to enact insurance legislation following the Panic of 1907, and was motivated partly by the adoption of an insurance system by Oklahoma in 1908.(13) In contrast to the Oklahoma system, in which all state-chartered banks were required to carry insurance, the Kansas system was made optional for state banks because of complaints that insurance forces conservatively managed banks to insure depositors of banks that are more likely to fail.(14) The state of Kansas, like other states with deposit insurance systems, did not guarantee insurance payments. In contrast to the experience with federal insurance in the 1980s, depositors, not taxpayers, suffered from any insurance fund deficiencies.(15)

Kansas banks were required to operate for at least one year and undergo an examination by state authorities before being admitted into the insurance system.(16) Insured banks were also required to maintain total capital of at least 10 percent of total deposits, and surplus and undistributed profits of at least 10 percent of total capital.(17) At first, deposit insurance was restricted to noninterest bearing accounts, savings deposits of $100 or less, and time deposits of between six and 12 months maturity. Banks with insured deposits were not permitted to pay more than 3 percent interest on any deposit, whether insured or not.(18) Regulations were relaxed in 1911; insurance was extended to all deposits not otherwise secured, including savings accounts in excess of $100, and the state banking commissioner was given authority to adjust interest rate ceilings as he deemed appropriate.

Insured banks were assessed annual premiums equal to 1120 of 1 percent of their insured deposits less total bank capital. Although a bank could reduce its premium by increasing its capital, the saving was small. A bank with $100,000 of insured deposits and $10,000 of capital -was assessed an insurance premium of $45, whereas a bank with $15,000 of capital had a premium of $42.50. Additional premiums could be assessed to cover shortfalls in the insurance fund, but total annual premiums were capped at 1/4 of 1 percent of insured deposits less capital. Banks also were required to place cash or eligible bonds with the state banking commissioner equal to 0.5 percent ($500 minimum) of their insured deposits to guarantee insurance premium payment. Banks could withdraw from the insurance system at any time, but remained liable for any premiums needed to reimburse depositors of banks which failed while the withdrawing bank was insured, including the six months following notice of withdrawal.

Deposit insurance proved popular in Kansas, and before 1920 the deposits of insured banks grew more rapidly than those of uninsured state and national banks. Figure 1 plots the participation rate of all Kansas banks and of those eligible for deposit insurance.(Figure 1 omitted) Figure 2 illustrates the shares of all bank and eligible bank deposits held by insured banks.(19)(Figure 2 omitted) The percentage of the state's bank deposits held by insured banks peaked in 1921 at 43.8 percent, and membership in the system peaked at 65.6 percent of eligible banks in 1923. In that year, 681 banks, holding $168 million of deposits, belonged to the insurance system, while 357 state banks, holding $64 million of deposits, did not.

THE CHARACTERISTICS OF INSURED BANKS

This section identifies some important differences between insured and uninsured banks that may explain why the failure rate of insured banks exceeded that of uninsured banks.

If depositors believe that they will be protected from loss in the event of bank failure, they will be willing to accept a lower rate of return on their deposits than they would in the absence of such protection. Because it lowers the cost of deposits, deposit insurance encourages banks to rely more heavily on deposits to **finance** their activities, as opposed to equity and nondeposit liabilities, than they otherwise would. Economic theory suggests that banks also will choose to hold riskier assets when deposits are insured.(20) Insured banks in Kansas had a higher failure rate than uninsured banks, which might have been caused by "moral hazard," that is, by high-risk behavior encouraged by deposit insurance. Alternatively, because risky banks would stand to gain the most from insurance in terms of lower deposit costs, the failure rate of insured banks might have been higher simply because risky banks were more likely to join the voluntary insurance system, that is, because of "adverse selection." Of course, both effects might have been present and contributed to the higher failure rate of insured banks.

The troubled history of the Kansas deposit insurance system raises the question of whether depositors expected an insurance payoff in the event of bank failure. If they did not, then depositors would have had an incentive to monitor their banks' activities and to demand the same terms from a member of the insurance system as from an uninsured bank with equal likelihood of failure. Indeed, if depositors thought that insured banks had, on average, a higher probability of failure and that an insurance payoff was unlikely, then they would have had an incentive to transfer deposits from insured banks to uninsured banks. No doubt some depositors did s4 as the relative share of deposits in insured banks fell after 1921. Large numbers of depositors left their funds in insured banks, however, and because of the difficulty of assessing the extent of protection from deposit insurance at any point in time, might have expected at least partial reimbursement in the event of bank failure.(21)

To investigate the relationship between deposit insurance and bank behavior, we compare various **financial** ratios of insured and uninsured banks in our sample in different years. Table 1 reports the mean capital/assets, deposits/assets and cash reserves/deposits ratios of insured and uninsured banks in our sample in each year for which data are available.(2)(Table 1 omitted) In general, insured banks maintained less capital relative to assets than uninsured banks and, hence, were more likely to fail as a result of loan losses or other declines in asset values. The hypothesis that the mean capital/assets ratios of insured and uninsured banks are equal can be rejected (at the .10 level or better) in each year.

The greater reliance of insured banks on deposits is indicated by the fact that, except for 1924, insured banks had higher deposits/ assets ratios than uninsured banks. Insured banks also held fewer liquid assets ("reserves"), defined here as cash, cash items and the liabilities of other banks, relative to deposits than uninsured banks in 1910, 1914 and 1924. Thus, for some of the period, insured banks were less liquid than their uninsured competitors. We find the reserves/deposits ratio to be particularly useful for distinguishing failing and nonfailing banks. The comparatively low capital/assets and reserves/deposits ratios of insured banks indicate that they were more risky than uninsured banks and, hence, the higher failure rate of insured banks is not surprising. We further examine the impact of deposit insurance on the probability of failure, and seek to identify other characteristics which distinguish failing from nonfailing banks in the following sections.

THE BANKING COLLAPSE OF THE 1920s

The number of banks and total bank deposits grew rapidly throughout the United States in the first two decades of the 20th century, especially during the inflationary boom of World War I. Kansas experienced a 30

percent increase in the number of banks between 1910 and 1420, when it had 1,096 state-chartered banks, 266 national banks and 18 unincorporated banks (Kansas, 1920, and Bankers Encyclopedia Company, March 1921). After 1920, the number of banks in the United States fell sharply, especially in the Midwest and the South, where waves of bank failures followed a collapse of commodity prices. Between June 1920 and January 1921, an index of wholesale commodity prices fell from 167 to 114; by January 1922, it had fallen to 91 (Board of Governors, 1937, p. 174). Sharply lower incomes left many farmers who had borrowed to **finance** land acquisition and improvements before 1920 unable to repay their loans. Loan losses, in turn, caused the failure of many banks in commodity-producing regions, including 220 in Kansas between 1920 and 1929.

The impact of agricultural distress on individual Kansas banks reflected the **portfolio** choices they had made prior to the collapse and as it unfolded. Between September 1920 and September 1926, 122 state-chartered Kansas banks failed. Of those, 94 had been members of the insurance system (a 4.6 percent failure rate) and 28 had not (a 2.3 percent failure rate). By contrast, just six national banks failed (a 0.8 percent failure rate). Over the life of the insurance system, depositors of just 27 failed banks recovered the entire amount of their insured deposits, and those of two other banks received 93 and 95 percent of their deposits, respectively (Warburton, 1958, pp. 27-9). No insurance payments were made to depositors of 88 member banks that failed (FDIC, 1956, p. 58). On average, holders of insured deposits received 53 percent of their funds from liquidation of bank assets and 18 percent from the deposit insurance fund (7 percent of which came from the reorganization of one bank, the American State Bank of Wichita). The remaining 29 percent of insured deposits were never recovered.

The sharp increase in bank failures beginning in 1920 quickly swamped the resources of the Kansas deposit insurance fund. When a member of the Kansas insurance system failed, its depositors were given interest-bearing certificates immediately upon closure, and received reimbursement only after the bank's assets had been entirely liquidated. If the proceeds from liquidation were insufficient to reimburse insured depositors, the insurance system was supposed to make up the difference. Depositors of the two banks that failed before 1920 were eventually fully reimbursed, but inadequate insurance funds meant that depositors of most banks that failed after 1920 were not as fortunate.

Because depositors were not reimbursed until after liquidation of a failed bank's assets, the condition of the fund and the prospect that depositors of failed banks would eventually receive full reimbursement were difficult to determine at any point. The failure in June 1923 of the American State Bank of Wichita, the state's largest insured bank, presented the insurance system with its greatest challenge. Eventually, the bank was reorganized with other insured banks assuming $1.4 million of the loss and depositors accepting, on average, 40 percent of their deposits in the form of stock in the new bank. The event marked a turning point in the history of the Kansas insurance system, however, as the number of banks and the deposits held in insured banks began to decline.(23)

Although a special insurance assessment was collected in 1922 and insurance premiums were set at their legal maximums beginning in 1924, losses from bank failures exceeded insurance system revenues from 1921 onward. In 1925, the state bank commissioner stopped making payments on all insurance claims, and in 1926 a state supreme court decision effectively ended the system. The court decision resulted from the refusal of several banks that had withdrawn from the insurance system to pay additional insurance premiums. The court ruled that banks could withdraw without additional liability by simply giving up the bonds they had pledged to guarantee premium payments. This led many banks to withdraw and, by 1927, insurance system membership had fallen to less than 20 percent of eligible banks.

Kansas appears to have suffered many of the problems that have been

associated with the bank and S&L debacle of the 1980s. In the 1980s, many depository institutions, especially insolvent S&Ls, bid up deposit interest rates and grew rapidly by issuing deposits through brokers.(24) In the 1920s, some banks appear to have evaded deposit interest rate ceilings in order to grow rapidly. In his report for 1922 (Kansas, 1992, p 5), the state bank commissioner also felt it desirable to limit deposit insurance to only the original holder of a deposit, and not to any assignee. Supervision was also reported to have been weak in Kansas, especially during the worst failure years, and for a time state banking authorities permitted weak and insolvent banks to remain open rather than closing them immediately upon recognition of trouble (Warburton, 1958, p. 19). Whether any such banks recovered is not known, but the lack of mention in the biennial reports of the state banking commissioner suggests that, like the attempts at forbearance during the 1980s, the policy was probably not successful.

MODELING TIME-TO-FAILURE

While many Kansas banks failed during the 1920s, a majority of banks survived the decade. What characteristics distinguish the survivors from the failures? identify important characteristics of failing banks, we employ an econometric technique that explicitly models time-to-failure. The **analysis** of duration data is relatively new in economics. Engineers and biomedical scientists have analyzed time-to-failure for electrical and mechanical components of machinery and the survival times of subjects for many years, but economists have only recently begun to apply similar models, primarily in the area of labor economics with a focus on the duration of spells of unemployment.(25) Although models developed to analyze duration data are sometimes called time-to-failure models, the event of interest need not be characterized as a "failure"; all that is necessary is that the event be well-defined.

Duration models differ from standard discrete choice models (such as probit or logit models) in that duration models use information about how long banks survive in the estimation of the instantaneous probability of failure for a given set of observations on the independent variables. Parameter estimates thus indicate whether an increase in the **value** of an individual independent variable will reduce or extend the expected time until failure occurs. By contrast, discrete choice models typically ignore information about the timing of failures, and provide an estimate only of the probability of failure within a given **interval** of time. Discrete choice models treat all banks that fail during an **interval** the same, as they do all surviving banks. Thus, for example, a bank that fails on the first day of a two-year **interval** is treated the same as a bank that fails on the last day, and a bank that survives the **interval** but fails one day after that period ends is treated the same as a bank that survives an additional 10 years. Duration models explicitly incorporate such information, and thus yield more efficient parameter estimates.(26) A detailed description of the duration model used in this article is presented in the appendix.

In the present application, we observe the charter date for each bank in our sample. For some banks, we observe a failure date, where failure is defined as the date on which the bank was ordered closed by the state banking commissioner. For the remaining banks, no failure date is observed if a bank had not failed by the end of our observation period (1928) or if it liquidated voluntarily, merged with another bank or switched to a federal charter. These observations are considered censored; information about these banks is available for part of their lives, but we do not observe them failing. Censoring is common in duration data of all types and must be addressed within the statistical model used to examine the data.

Figure 3 illustrates the types of censoring that may occur in duration data. Assume that the **interval** over which banks are observed runs from time t sub 1 to t sub 2 .(Figure 3 omitted) The horizontal lines in the figure represent the time between the charter date and the date of

failure for individual banks. Given the observation period (t sub 1 , t sub 2), the observation for Bank A will be both left- and right-censored. For this bank, neither the charter date nor the failure date occur within the observation **interval**. The observation for Bank B will be left-censored; the charter date does not occur within the observation **interval**, but the failure date does. For Bank C, both the charter and failure dates occur between t sub 1 , and t sub 2 , and so the observation is uncensored. Finally, the observation for Bank D will be right-censored; the charter date occurs within the observation **interval**, but the failure date occurs after t sub 2 .

EXPLANATORY VARIABLES AND ESTIMATION RESULTS

Other researchers have employed hazard and discrete choice models to identify characteristics that distinguish failing and surviving banks in a variety of settings. White (1984), for example, estimates a probit model to distinguish failing from nonfailing banks during the Banking Panic of 1930. Wheelock (1992a) uses a similar model to study Kansas bank failures between 1920 and 1926. Both studies found that banks were more likely to fail, the lower their capital/assets, surplus/loans, bonds/assets, reserves/deposits, or deposits/assets ratios.(27) Banks were more likely to fail, the higher their loans/assets or short-term borrowed funds/assets ratios.(28)

Many Kansas banks experienced significant loan losses following the collapse of agricultural prices and incomes in 1920-21, and banks with low capital/assets ratios were less well-cushioned against declines in the **value**lt; of their assets. Banks with little cash and other reserve assets were less able to meet deposit withdrawals, and the smaller a bank's reserves/deposits ratio the more likely it was to close due to illiquidity. Often a lack of cash was the first sign that a bank was in trouble, and would prompt closure by state banking authorities.

Just as a low level of reserves signaled trouble, so too did a heavy reliance on borrowed funds such as rediscounts of loans with other banks or with the Federal Reserve. Banks that relied heavily on borrowed funds to **finance**lt; their operations, or that had to resort to borrowing because of loan losses or deposit withdrawals, appear to have been relatively more likely to fail.

Loans are generally the most risky and least liquid of bank assets, and the loan portfolios of the rural unit banks of Kansas were undoubtedly not well-diversified. Accordingly, the higher a bank's loans/assets ratio, the greater the likelihood that it would fail. On the other hand, banks with substantial bond holdings might have been less likely to fail, especially since U.S. Government bonds and bonds of the state of Kansas and of Kansas municipalities probably comprised most of the bond holdings of Kansas banks in this era.(29)

Wheelock (1992b) includes bank size and a dummy variable indicating whether or not a bank was a member of the state deposit insurance system as additional explanatory variables. If larger banks were better diversified, or could capture economies of scale, they might have been less likely to fail. Wheelock found, however, no significant relationship between size and failure. Deposit insurance, on the other hand, did significantly affect the probability of failure. Even though the capital/assets ratio and other measures of risk-taking should reflect whether or not a bank had insured deposits, the full impact of insurance may not be captured by observable variables. The deposit insurance dummy variable might reflect the incentive that insurance gives banks to hold riskier loans and investments than they otherwise would. Wheelock (1992b) did not test for interaction effects between deposit insurance and the **financial**lt; ratios. One might expect, however, that the effect of a change in a **financial**lt; ratio on the likelihood of failure would depend in part on whether or not the bank had deposit insurance. For example, the depositors of an insured bank might have been less concerned with a decline in the capital/assets ratio of their bank and, hence, less likely to demand a higher deposit interest rate than depositors of an uninsured bank. The scope for risk-taking and, thus, the probability of failure, resulting from a change

in a **financial**lt; ratio might therefore depend on whether or not a bank was insured. We test this hypothesis here.

ESTIMATION RESULTS

Our data consist of a panel of Kansas banks for which we have collected balance sheets and other information as of August 31 of each even. numbered year from 1910 to 1926 (except 1912 and 1916, when these data were not published).(30) Our sample includes 259 banks (approximately one-fourth the total operating in 1914).(31) Of these, 47 (18 percent) had failed by September 1, 1928. Banks that merged with other banks, liquidated voluntarily or switched to a federal charter are treated as censored on the date of merger or change in charter. Banks that did not fail or otherwise ceased operating prior to August 31, 1928, are treated as censored on that date.

In addition to the independent variables used by Wheelock (1992b), we include dummy variables for each **interval**lt; of 1920-22, 1922-24, 1924-26 and 1926-28 to investigate whether the probability of failure differed across periods for a given set of bank attributes. Only two banks in our sample failed before 1920 and, hence, we do not include dummies for those years. In one specification we also include interaction terms of deposit insurance and the **financial**lt; ratios.(22)

Table 2 reports estimates of the failure model that include alternative combinations of explanatory variables.(Table 2 omitted) In column one, the coefficient on the capital/assets, bonds/assets and reserves/ deposits ratios indicate that the higher each of these ratios was, the less likely a bank was to fail. Better capitalized banks, and those with substantial bond holdings and significant reserves, could better absorb the shock of loan losses and deposit withdrawals accompanying the agricultural downturn in Kansas. Banks that had substantial borrowed funds relative to assets had a greater chance of failing while, contrary to expectations, it appears that the higher a bank's loan/assets ratio the less likely it was to fail. This finding appears due to multicollinearity, however. The loans/assets ratio is highly correlated with the reserves/deposits ratio. If the latter is omitted, as in the specification reported in column two, the sign of the coefficient on the loans/assets ratio is positive, though not statistically significant.

The coefficient on deposit insurance is not statistically significant, suggesting that any effect that insurance had on the probability of failure is captured by its relationship with the **financial**lt; ratios also included in the model. If the dummies for the biennial observation **intervals**lt; are omitted, the coefficient on insurance is larger and statistically significant. It may be that the strain on the portfolios of all banks caused by the collapse of commodity prices overwhelmed the effect of deposit insurance on the unobserved **portfolio**lt; risk of insured banks, which could explain why the coefficient on insurance is not significant when the time dummies are included. Not surprisingly, for given values of the **financial**lt; ratios, banks were more likely to fail after the collapse of commodity prices and onset of severe agricultural distress in 1920. Finally, none of the coefficients on the interaction terms of deposit insurance and the **financial**lt; ratios is statistically significant. Again, it appears that any impact of deposit insurance on the likelihood of failure is captured by differences in the **financial** ratios between insured and uninsured banks.

CONCLUSION

Researchers have blamed federal deposit insurance for contributing to the high numbers of bank and thrift failures and large deposit insurance payoffs since 1980. Unless insurance premiums increase proportionately with risk, banks will be encouraged to take greater risks than they otherwise would. This article presents some **historical** evidence of how deposit insurance can alter bank behavior and increase the likelihood that a bank will fail. As in the 1980s, when falling incomes in agricultural and energy-producing states caused large loan losses and led to many bank and thrift failures, a sharp decline in agricultural incomes in the early 1920s

caused the failure of many commercial banks in rural areas. Not all banks failed, however; in fact, most survived the collapse. Banks that carried deposit insurance had a higher rate of failure than other banks. Our findings, along with those of similar **historical** studies, show that insured banks were less well capitalized and less liquid than other banks. Estimates of a model of time-to-failure indicate that among banks in our sample, those with high ratios of capital to assets, reserves to deposits, large bond holdings relative to their total assets, or that relied little on borrowed funds, were less likely to fail. In short, conservatively managed banks were less likely to fail and, at the same time, banks that carried deposit insurance were more risky and, hence, more likely to fail than their uninsured competitors.

REFERENCES

Alston, Lee J., Wayne A. Grove and David C. Wheelock. "Why Do Banks Fail? Evidence from the 1920s," Explorations in Economic History (1994, forthcoming).

American Bankers Association. The Guaranty of Bank Deposits (1933).

Anderson, Per K., and Richard D. Gill. "Cox's Regression Model for Counting Processes," Annals of Statistics (December 1982), pp. 1100-20.

Association of Reserve City Bankers. The Guaranty of Bank Deposits (1933).

Bankers Encyclopedia Company. Polk's Bankers Encyclopedia (March 1921).

Board of Governors of the Federal Reserve System. All Bank Statistics, 1896-1955(1959).

--. Banking and Monetary Statistics, 1914-1941 (1943).

--. Annual Report (1937).

Calomiris, Charles W. "Deposit Insurance: Lessons from the Record," Federal Reserve Bank of Chicago Economic Perspectives (May/June 1989), pp. 10-30.

Cooke, Thornton. "The Insurance of Bank Deposits in the West," Quarterly Journal of Economics (Nov. 1909), pp. 85-108.

Cox, David R. "Partial Likelihood," Biometrika (May/August, 1975), pp. 269-276.

--. "Regression Models and Life Tables," Journal of the Royal Statistical Association (1972), pp. 187-220.

Efron, Bradley. "The Efficiency of Cox's Likelihood Function For Censored Data," Journal of the American Statistical Association (September 1977), pp. 557-565.

Federal Deposit Insurance Corporation. Annual Report (1956).

Flood, Mark D. "The Great Deposit Insurance Debate," this Review (July/August 1992), pp. 51-77.

Golembe, Carter. "The Deposit Insurance Legislation of 1933: An Examination of its Antecedents and its Purposes," Political Science Quarterly (June 1960), pp. 181-200.

Grossman, Richard S. "Deposit Insurance, Regulation and Moral Hazard in the Thrift Industry: Evidence from the 1930s," The American Economic Review (September 1992), pp. 800-21.

Harger, Charles Moreau. "An Experiment that Failed: What Kansas Learned by Trying to Guarantee Bank Deposits," Outlook (June 23, 1926), pp. 278-79.

Johansen, Soren. "An Extension of Cox's Regression Model," International Statistical Review (August 1983), pp. 165-74.

Kalbfleisch, J. D. and R. L. Prentice. The Statistical **Analysis** of Failure Time Data. John Wiley and Sons, 1980.

Kane, Edward J. The S&L Insurance Mess: How Did It Happen? The Urban Institute Press, 1989.

Kansas. Biennial Report of the Bank Commissioner (various years).

Kareken, John H., and Neil Wallace. "Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition," Journal of Business (July 1978), pp. 413-38.

Keeley, Michael C. "Deposit Insurance, Risk, and Market Power in

Banking," The American Economic Review (December 1990), pp. 1183-1200.
Kiefer, Nicholas M. "Economic Duration Data and Hazard Functions," Journal of Economic Literature (June 1988), pp. 646-79.
Lancaster, Tony, The **Analysis** of Transition Data. Cambridge University Press, Inc., 1990.
Merton, Robert C. "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees," Journal of Banking and **Finance** (June 19/7), pp. 3-11.
Mishkin, Frederic S. "An Evaluation of the Treasury Plan for Banking Reform," Journal of Economic Perspectives (Winter 1992), pp. 133-53.
O'Driscoll, Gerald P. "Bank Failures: The Deposit Insurance Connection," Contemporary Policy Issues (April 1988), pp. 1-12.
Robb, Thomas B. The Guaranty of Bank Deposits. Houghton Mifflin Co., 1921.
Warburton, Clark. Deposit Guaranty in Kansas, unpublished manuscript, Federal Deposit Insurance Corporation, 1958.
Wheelock, David C. "Is the Banking Industry in Decline? Recent Trends and Future Prospects From a **Historical** Perspective," this Review (September/October 1993), pp. 3-22.
--. "Regulation and Bank Failures: New Evidence from the Agricultural Collapse of the 1920s," The Journal of Economic History (December 1992a), pp. 806-25.
--. "Deposit Insurance and Bank Failures: New Evidence from the 1920s," Economic Inquiry (July 1992b), pp. 530-43.
--, and Paul W. Wilson. "Explaining Bank Failure: Deposit Insurance, Regulation and Efficiency," working paper no. 93-002A. Federal Reserve Bank of St. Louis (1993).
--, and Subal C. Kumbhakar. "'The Slack Banker Dances': Deposit Insurance and Risk-Taking in the Banking Collapse of the 1920s," Explorations in Economic History (1994 forthcoming).
White. Eugene N. "A Reinterpretation of the Banking Crisis of 1930," Journal of Economic History(March 1984), pp. 119-38.

APPENDIX THE PROPORTIONAL HAZARD MODEL

This appendix describes the proportional hazard model estimated in this article in some detail for interested readers who are unfamiliar with duration models, but have some understanding of econometrics or statistics. We wish to estimate the effect of deposit insurance and other variables on the probability of failure at particular times for the n banks in our sample. Let T sub i , i = 1,..., n, represent the failure time for the ith bank in our sample; where T sub i is not observed, we say that the observation is censored. Time is measured relative to the individual bank's charter date, with a zero **value** at the charter date. Hence, for each bank in the sample, the corresponding time scales will have different values for a given calendar time. If T is a continuous **random** variable with a continuous probability distribution f(t), where t is a realization of T, then the cumulative probability is

(Equation 1 omitted)

The function F(t) gives the probability that a bank fails before time t (subscripts are omitted where no ambiguity results). Alternatively, the same information may be expressed in terms of the survival function

(2) S(t) = 1 - F(t),

which is merely Prob(T>/-t).

Given that a bank has survived until time t, what is the probability that it will fail during the next short **interval** of time, M The function characterizing this aspect of the problem is the hazard rate, given by

(Equation 3 omitted)

The hazard function gives the instantaneous rate of failure per unit time period at time t. The density function f(t), the cumulative density function F(t), the survival function S(t) and the hazard function A(t) each characterize the duration of banks. Furthermore, all four functions are related. From equations 2 and 3, it follows that

(Equation 4 omitted)

Also rearranging terms in equation 3 yields

(5) t) = S(t)lambda(t).

Another useful function is the integrated hazard function,

(6) (Equation 6 omitted)

Then, from equations 4 and 6 the survival function may be written as

(Equation 7 omitted)

and from equation 7 we have

(Equation 8 omitted)

Estimation of the failure time relationship requires specifying a functional form for either f(t), F(t), S(t), lambda(t), or (t). Note that a functional form need only be specified for one of these functions; the relations in 1 and 4-8 will imply a functional form for the remaining functions.

We use the proportional hazard relationship developed by Cox (1972h where

(Equation 9 omitted)

where is a row vector of measured covariates and beta is a column vector of parameters with the appropriate dimensions. This model assumes a baseline hazard, lambda sub o ,(t), which in principle amounts to an unidentified parameter for each bank in the sample. Thus, lambda sub o (t) accounts for any unobserved heterogeneity among the banks that might otherwise bias the parameter estimates. The covariates in chi influence the overall hazard for each bank through the exponential terms in equation 9 (the choice of an exponential form here is common throughout the literature on hazard estimation and simplifies the estimation problem relative to choices of other functional forms). The model is semiparametric since the exponential in 9 is a parametric form, while the baseline hazard involves an unspecified form and, hence, is nonparametric. Consequently, the model is more flexible than models in which the failure time distribution is assumed known except perhaps for a few scalar parameters.

Given the hazard specification in equation 9, the corresponding survivor function (which gives the probability of survival up to time t) may be written as

(Equation 10 omitted)

For uncensored observations with failure at time the contribution to the likelihood is f(T chi); for observations censored at time T, the contribution to the likelihood is S(T chi), that is, the probability of survival until time

Cox (1972, 1975) suggests a partial-likelihood approach which can be used to estimate the parameters of the hazard function in 9. Assume, for the moment, that no observations are censored, and that the observations are ordered by their completed, untied durations such that t, < t sub 2 < ... < t sub n . The conditional probability that observation 1 fails at time t sub 1 , given that any of the n observations could have failed at time t sub 1 , is

(Equation 11 omitted)

The equality results from the assumption of the proportional hazard in 9; the baseline hazard lambda sub o (t) cancels out of the expression on the left in 11. The expression in 11 gives the contribution of the first observation to the partial likelihood. Analogously, the contribution of the jth observation to the partial likelihood is

(Equation omitted)

The partial likelihood is given by the product of the individual contributions and, hence, its log is

(Equation 12 omitted)

Andersen and Gill (1982) and Johansen (1983) show that the partial likelihood can be treated as an ordinary likelihood concentrated with respect to lambda sub o .(33)

The model represented by equation 12 can be easily adjusted to accommodate censoring in the data. For the data used in this study, each bank i in the sample is observed at J different times t sub i1 < t sub

i2 < ... < t sub ij , with either failure or censoring occurring at time t sub ij . Note that times here refer not to calendar time, but to time relative to the date of charter for bank i so that t sub io , = O where t, is the date of charter for the ith bank. The balance sheet information used in corresponding to time t sub ij , j = 1, ..., (J sub i -1), are assumed to reflect the position of bank i over the **interval** (t sub ij , t sub i(j+1) .) The model estimated in this paper is time-varying in the sense that covariates in are assumed constant for **intervals** of time (t sub ij , t sub i(j+1)), but may vary across different **intervals**. Thus, for the ith bank there are (J sub i -1) observed **intervals**; the first (J sub i -2) are both left-and right-censored, and the last is left-censored and also right-censored if failure time is not observed for the ith bank.

To accommodate the censoring in the data, let chi sub i (J), i=1, ..., n, j=1, ..., J sub i denote the vector of covariates for bank i during period j. Covariates are fixed within a given period, but may vary over different periods. Let d sub i equal 1 for banks that are observed to fail at some time within the entire observation period, and zero otherwise. Assume that banks are ordered by increasing date of failure. Then, the log-partial likelihood becomes

(Equation 13 omitted)

Kiefer (1988) suggests that the intuition behind the partial-likelihood approach used here is that, in the absence of any information about the baseline hazard, only the order of the durations provides information about the unknown parameters of the model. In both 12 and 13, the instantaneous probability of failure is normalized by the sum of instantaneous probabilities of failure for all other banks that could have failed at the same time as the ith bank.

1 Kane (1989) examines the problems of the S&L industry and the role of government policy. Mishkin (1992), Keeley (1990) and O'Driscoll (1988) discuss the relationship between deposit insurance and bank failures in the 1980s.

2 Wheelock (1992b) also investigates how deposit insurance affected the probability of failure for Kansas banks in this era, but employs a different methodology and somewhat different data. Wheelock and Wilson (1993) use the same data set as the present study, but while considering the effects of insurance, focus largely on whether or not measures of managerial inefficiency help distinguish tailing from surviving banks. Grossman (1992) also investigates the effects of deposit insurance by comparing insured and uninsured S&Ls during the 1930s.

3 Studies of the causes and effects of bank failures during the Depression are too numerous to list. Friedman and Schwartz (1963), however, is the seminal investigation of the impact of bank failures on the money supply, and Bernanke (1983) is the most important investigation of nonmonetary effects of bank failures.

4 Golembe (1960) and Flood (1992) investigate the rationale for federal deposit insurance.

5 For example, see Friedman and Schwartz (1963, pp. 434-42).

6 For example, see Kane (1989).

7 Keeley (1990) draws the connection between increased competition, deposit insurance and increased risk-taking.

8 Too-big-to-fail was implemented to reduce the possibility that the failure of a very large bank could produce a systemic crisis, with depositor runs on many banks. Mishkin (1992) and Boyd and Gertler (1993) argus that this policy increased risk-taking by very large banks.

9 See Wheelock (1993).

10 The 20th-century states and the years in which their insurance systems operated are Oklahoma (1907-23), Texas (1909-25), Kansas 909-29), Nebraska (1909-30), South Dakota (1909-31), North Dakota (1917-29), Washington (1917-29) and Mississippi (1914-30). Cooke (1909), Robb (1921), American Bankers Association (1933), Federal Deposit Insurance Corporation (1956) and Calomiris (1989) compare the features and performance of the

systems.

11 Thies and Gerlowski (1989) and Alston, Grove and Wheelock (1994) find that a state's bank failure rate during the 1920s was higher if it had a system of deposit insurance, holding constant other possible causes of failure. Wheelock (1992a) reports similar evidence at the county level for Kansas.

12 Commenting about the effects of the Kansas deposit insurance system, Harger (1926, p. 278) wrote that insurance "gave the banker with little experience and careless methods equality with the manager of a strong and conservative institution. Serene in the confidence that they could not lose, depositors trusted in the guaranteed bank. With increased deposits, the bank extended its loans freely See also American Bankers Association (1933), Association of Reserve City Bankers (1933) and Robb (1921) for contemporary views about insurance.

13 Robb (1921) describes previous attempts to enact deposit insurance legislation in Kansas and other states, and notes that Kansas banks located near the Oklahoma border were especially strong proponents of deposit insurance in Kansas (pp. 107-12).

14 The Comptroller of the Currency ruled in 1908 that national banks could not join state deposit insurance systems.

15 Mississippi, however, ultimately issued bonds to retire the deficit of its insurance system.

16 The requirement of one year of operation was waived if no other bank in the applicant's town was an insurance system member.

17 The former requirement was eliminated in 1917. Warburton (1958, p. 21) argues that, if maintained and enforced, the requirement could have prevented much of the rapid growth of banks that ultimately resulted in large losses to the insurance system.

18 For comparison, the annual average interest rates on prime four-six month commercial paper and on call loans in 1909 were 4.69 and 2.71 percent, respectively.

19 All banks include those with federal charters, trust companies and unincorporated banks. The source of these data is the FDIC (1956, p. 68).

20 See Merton (1977) or Kareken and Wallace (7978).

21 Wheelock and Kumbhakar (1994) argue that before 1926, depositors had a reasonable expectation of an insurance payoff, and show that deposit insurance enabled members of the insurance system to hold lower capital ratios than uninsured banks until that year.

22 The biennial reports of the state banking commissioner (Kansas, various years) provide balance sheet data for all state-chartered banks and trust companies on August 31 of each even-numbered year (except 1912 and 1916).

23 See Wheelock and Kumbhakar (1994) and Warburton (1958) for additional detail about this failure.

24 See Kane (1989).

25 Kiefer (1988) provides a good introduction to the **analysis** of duration data; Kalbfleisch and Prentice (1980) and Lancaster (1990) provide more advanced treatments of the subject.

26 While deriving a direct relationship between the parameters of a duration model and a discrete choice model would be difficult, in principle one could integrate the hazard function estimated from a duration model to obtain the probability of failure within a given **interval** of time.

27 Surplus refers to paid-in capital beyond the par **value** of a bank's stock plus undistributed profits. Reserves refer to cash, cash items and the liabilities of other banks.

28 Borrowed funds consisted largely of rediscounted loans with the Federal Reserve or other banks.

29 The state banking commissioner accepted only U.S. Government, state of Kansas and Kansas municipal bonds to guarantee payment of deposit insurance premiums. Unfortunately, we do not have information on the composition of each bank's bond holdings.

30 The source of our data is Kansas (various years).

31 We dropped seven banks because of missing data. Others fall out of the panel after failing, closing voluntarily, merging with other banks, or switching to a national charter.

32 Few state-chartered Kansas banks were members of the Federal Reserve System during this era. None of the failed banks in our sample was a member, and so differences in supervisory agency or regulation, except those pertaining to deposit insurance, cannot explain variation in failure probabilities across banks.

31 Alternatively, one could specify a parametric form for the baseline hazard in equation 9 and maximize the corresponding likelihood function. Although the partial-likelihood approach avoids the need for an arbitrary parametric specification of the baseline hazard, there is a loss of efficiency in the resulting estimates relative to those obtained by maximizing the full likelihood. See Efron (1977) for a discussion of this efficiency loss.

Special Features: Charts; Graphs; Diagrams; Appendix; References
Classification: 4310 (CN=Regulation); 8100 (CN=Financial services industry); 9190 (CN=United States)
Descriptors: Deposit insurance ; Impacts; Bank failures; Risk
Geographic Names: US

36/9/21 (Item 21 from file: 268)
00242836
The role of large banks in the recent U.S. banking crisis
Boyd, John H; Gertler, Mark
Federal Reserve Bank of Minneapolis Quarterly Review , v 18 , n 1 , p 2-21 , Winter 1994 **Document Type:** Journal Article **ISSN:** 0271-5287 **Journal Code:** FMQ **Language:** English **Record Type:** Abstract Fulltext
ARTICLE REFERENCE NUMBER:
Word Count: 10326

Abstract:

In the 1980s, US commercial banking faced its greatest crisis since the Great Depression. Because of increased competition from open market sources of credit and nonbank intermediation and a series of shocks to loan portfolios, banks saw shrinking profits and a growing likelihood of failure. By the end of 1992, the FDIC listed 863 banks with combined assets of $464 billion as problem institutions. It is shown that banks with the largest total assets contributed in a significantly disproportionate way to aggregate loan losses. Two factors allowed this to happen: 1. Deregulation and **financial** innovation led to increased competition. 2. The existing regulatory environment tended to subsidize risk taking by large banks more than that by small banks. Under a policy known as "too-big-to-fail," large banks benefited from a multitude of actions that insulated them from the impact of their loan losses. This policy was a key factor behind the recent crisis.

Text:

The 1980s witnessed the greatest crisis in U.S. commercial banking since the Great Depression. Faced with both increased competition from open market sources of credit and nonbank intermediation and a series of adverse shocks to loan portfolios, banks experienced shrinking profits and a growing likelihood of failure. Indeed, the failure rate for banks jumped from an average of 2 per year in the 1970s to roughly 130 per year in the period between 1982 and 1991. Accompanying the increase in the failure rate was a rise in the number of banks in **financial** distress. By the end of 1992, the Federal Deposit Insurance Corporation (FDIC) listed 863 banks with combined assets of $464 billion as problem institutions (FDIC 1993, p. 7).

It is true that the situation has improved very recently. Low short-term interest rates have permitted banks to widen the spread between deposit and loan rates and to exploit a favorable yield curve. As a consequence, profits have risen over the last several years. Nonetheless, it is still important to examine what went wrong during the 1980s, to identify what lessons can be learned to guide regulatory policy.

In this article, we argue that large banks were mainly responsible for the unusually poor performance of the overall industry. As we show, banks with the largest total assets contributed in a significantly disproportionate way to aggregate loan losses. Two factors allowed this to happen. First, deregulation and **financial** innovation led to increased overall competition for the banking industry. Second, the existing regulatory environment tended to subsidize risk-taking by large banks more than that by small banks. Under the policy known as too-big-to-fail, large banks benefited from a multitude of actions that insulated them (to varying degrees) from the impact of their loan losses. These actions ranged from giving large banks favorable treatment at the discount window to actually providing them with direct subsidies to prevent their failure. (Thus, by too-big-to-fail we refer to a menu of policies. And it is likely that only the very largest banks had access to the complete menu.) The rationale behind too-big-to-fail was that, based on the experience of the Great Depression, the failure of a large bank could be contagious. It could greatly disturb the rest of the **financial** system and cause severe consequences for the entire U.S. economy. But this well-intentioned policy had an unfortunate side effect: it unduly subsidized risk-taking by large banks.

Our view that the too-big-to-fail policy was a key factor behind the recent crisis comes from several pieces of information. One is a panel data study that we use to sort out the potential sources of loan losses. It is well known that regional economic factors initiated a significant share of loan losses during the 1980s. For example, collapsing oil prices in the Southwest and real estate prices on the East and West Coasts produced regional banking crises during this time. So it may be possible that the relatively poor performance of large banks could be explained by the fact that these banks tended to be clustered in the hard-hit regions. We show, however, that after regional conditions are controlled for, size still matters in explaining loan losses. That is, on average, even after the influence of region is moved from the data, large banks still performed worse than other banks. As we show, this conclusion is robust to a very general specification of how size and region may interact to influence bank performance.

Our second piece of evidence comes from the composition of portfolios across size classes. We find that in the 1980s large banks tended to operate with a much lower buffer against potential loan losses. In particular, they held unusually low ratios of capital to assets relative to

the industry mean. Up to a certain asset size, a negative relationship between the capital/assets ratio and size might be explained by diversification gains and increased access to purchased money markets as a bank grows larger.(1) However, we find that the capital/assets ratio tended to decline markedly with size well beyond the point that might naturally be explained by these scale economy factors. In particular, the capital/assets ratio shrank significantly even after banks moved above $10 billion in total assets. To our knowledge, there is no evidence of the scale economies we have described once banks moved above this size range. In addition to the capital/assets ratio, we consider several other measures of **portfolio** risk and similarly find that large banks adopted a riskier stance, beyond what could sensibly be explained by scale economies.

In the first section below, we present information on the recent trends in banking, in order to provide a sense of context. We describe the evolution of the combination of factors, including the development of the too-big-to-fail policy, that provided the climate for the subsequent crisis. We also present evidence on the composition of bank portfolios across size classes.

In the next section of the article, we first present measures of bank performance during the 1980s crisis and then present the results of the panel data study. We find that loan losses varied significantly with size, even after we control for regional conditions. We also use our results to compute a rough estimate of the impact of large banks' extranormal loss performance on the industry aggregate and find that this impact was quite substantial. In addition, we present evidence that it was mainly large banks that were constrained during what some call the capital crunch of the early 1990s. Thus, though investigating the importance of the capital crunch is clearly beyond the scope of this study, we provide insights into the origin of this phenomenon.(2)

In the final section of the article, we summarize our study and offer some thoughts on related issues. While we think that the regulatory subsidy to large banks has been a significant factor behind the recent banking crisis, we should make it clear that we are not advocating any kind of sweeping withdrawal of the safety net. Indeed, as we discuss, some of the recent banking reforms are steps toward addressing the too-big-to-fail distortion. We are, however, skeptical about the benefits of mergers that create even larger banks.

DOCUMENTING BALANCE SHEET TRENDS

We begin by describing the changes in U.S. banking over the postwar period that laid the groundwork for the problems of the 1980s. We document the general shift in bank portfolios in favor of riskier asset and liability positions, and we show how the too-big-to-fail policy evolved in this environment. We also show that the shift in favor of increased balance sheet risk was particularly characteristic of large banks.

Banks Take Bigger Risks

* Assets

Judged by a variety of criteria, the composition of bank assets has become riskier over the postwar period.

Chart 1 portrays the relative behavior of the broad categories of bank assets over the postwar period. (Chart 1 omitted) Most striking are the rise in the share allocated to loans and the fall in the shares allocated to securities and to cash and reserves. The drop in the latter reflects mainly a sequence of reductions in reserve requirements. An important reason for the secular (or long-run) decline in the security share is the development of money markets, such as the federal funds and large certificate of deposit (CD) markets. The increased access to short-term money permitted banks to reduce precautionary holdings of securities. Also, certain types of bank loans became increasingly liquid over time due to the advent of securitization and the development of markets for loan sales. Recently, the share of securities has been rising--partly due to the problems in banking and the associated regulatory changes and partly due to banks exploiting the steepness of the yield curve.

Chart 2 disaggregates bank loans. (Chart 2 omitted) The main categories are commercial and industrial (C&I) loans, mortgages, and consumer credit. Interestingly, the shares of each in bank loan portfolios were relatively stable from 1952 to about 1973. Since then, though, the share of C&I loans has declined, and since the early 1980s, the decline has been fairly precipitous. One factor underlying this trend has been the growth of the commercial paper market, which largely involved a movement of high-quality C&I lending off bank balance sheets. Another factor is the growth of nonbank intermediation, particularly **finance** company lending, as Chart 3 illustrates. (Chart 3 omitted)

A less well-known factor underlying the relative decline in C&I loans is the recent growth of offshore commercial lending. While the Federal Reserve's flow of funds measure of C&I lending includes commercial lending both by domestic banks and by branches of foreign banks within the U.S. border, this measure underestimates loans to U.S. firms by banks located offshore. The market for offshore lending grew rapidly during the 1980s. One likely reason for this rapid growth, according to McCauley and Seth (1992), is that differences in reserve requirements on large CDs made intermediating (high-quality) loans cheaper offshore. Banks lending onshore were required to hold 3 percent reserves against large CDs; offshore banks faced no reserve requirements. Chart 4 illustrates the growth of offshore commercial loans. (Chart 4 omitted) These loans grew from 7 percent of total C&I lending in the United States in 1983 to more than 20 percent by 1991. (For a detailed **analysis** of this somewhat surprisingly large change, see McCauley and Seth 1992.) Further, since the offshore banks are at some disadvantage in the evaluation and monitoring of small and medium-sized companies, the type of commercial loan business they absorbed was likely the loans to larger, better-rated companies.(3)

While high-quality commercial lending moved off bank balance sheets to both domestic and foreign competition, the relative importance of mortgage lending grew. This phenomenon began in the mid-197Os and accelerated throughout the 1980s. (See Chart 2.) Banks undoubtedly picked up some business from failing savings and loan associations (S&Ls), especially in the latter half of the 1980s. However, the shift to mortgage lending occurred well prior to the S&L debacle.

Disaggregating mortgage lending uncovers another important trend. As Chart 5 shows, commercial mortgage lending has accounted for much of the recent growth in overall bank mortgage lending.(4) (Chart 5 omitted) In 1980, home mortgages accounted for about 60 percent of bank mortgage lending, and commercial mortgages accounted for about 30 percent. By 1990, the shares of the two types were about equal, each roughly 45 percent of overall bank lending.(5) This phenomenon is of interest since many of the problems in banking stem from losses in commercial real estate lending, as we discuss later. In this context, it is important to note that the marked shift of banks from residential to commercial mortgages was not symptomatic of mortgage lending in general. Chart 6 shows that, for all **financial** intermediaries, the shares of aggregate mortgage lending going to the residential and commercial sectors have been relatively stable. (Chart 6 omitted)

The movement of banks into commercial real estate reflects part of a broader trend in bank lending since the 1970s. High-quality assets such as securitized residential mortgages or commercial loans to highly rated firms have moved off bank balance sheets. In a fight to maintain market share, banks have exploited their comparative advantage in information-intensive lending by moving into riskier, less liquid assets. Banks' comparative advantage stems partly from experience in evaluating and monitoring. It also stems partly from the nature of the regulatory system, particularly the nature of the public safety net. Later we will return to these issues.

* Liabilities

Judged by a variety of criteria, the composition of bank liabilities also appears to have become riskier.

The flow of funds accounts divide bank liabilities into four categories: checkable deposits, small time and savings deposits, money market liabilities, and long-term debt. Chart 7 shows the long-term trends. We see two important patterns in the data. (Chart 7 omitted)

Perhaps the most obvious pattern is the secular decline in the relative importance of checkable deposits, in favor of interest-bearing liabilities.(6) As late as 1960, nearly 60 percent of bank liabilities were checkable deposits and only about 30 percent were small time and savings deposits. The use of money market instruments and long-term debt was negligible. By 1990, checkable deposits were least important, less than 20 percent of total liabilities. Small time and savings deposits had climbed to 40 percent, while money market instruments and long-term debt each had climbed to around 20 percent.

The second important pattern, which is closely related to the first, is the increased use of managed liabilities relative to deposits. Managed liabilities are short-term instruments which pay market-determined rates of interest. In contrast to deposits, which a-e relatively immobile in the short run, managed liabilities are highly inter-elastic. Managed liabilities permit banks to rapidly adjust their stock of loanable funds. Money market instruments are the prime example. There are two main types of money market liabilities (also known as purchased money): large time deposits and federal funds plus security repurchase agreements. The former (large CDs) typically have maturities that vary from 90 days to a year, while the latter consist largely of overnight and weekly loans. The use of both types of instruments grew sharply in the early 1970s, as deregulation permitted the development of the money market.

Recently, banks also appear to have been treating small CDs as managed liabilities. With deregulation of rates, small CDs have become increasingly sensitive to market forces. About two-thirds of small time and savings deposits are small CDs. Thus, if we include small CDs along with money market instruments in the measure of managed liabilities, these liabilities now constitute more than half of banks' short-term obligations.

The increased use of managed liabilities--and of money market instruments in particular--has had a number of important effects. One obvious effect is downward pressure on banks' net interest margins (the difference between the return per dollar on the asset **portfolio** and the interest cost per dollar of liabilities). Another effect is a rise in the interest sensitivity of bank liabilities. Today, in contrast to years past, an adverse movement in short-term rates may substantially raise banks' interest expenses. The development of the money market has also reduced the constraints of restrictions on interstate banking. The money market permits banks to cross state borders (or in the case of the money center banks, to cross international borders) in order to obtain short-term funds.

It is also true that the development of the money market has posed a vexing problem for regulators. In some ways, the failure of the regulatory system to appropriately adapt to the changes introduced by the money market was the precursor to the problems the banking industry faces today. With the efficiency gains of the money market came the cost of increased exposure to liquidity risk. While textbook descriptions of bank runs still conjure up images of people rushing through the doors of depository institutions with passbooks in hand, the most likely source of a widespread banking collapse today is a panic withdrawal of money market instruments. Since these instruments typically have values in excess of $100,000, they are not covered by deposit insurance. For this reason, and because they are highly mobile funds, abrupt withdrawal is a possibility. The key point is that, in the current environment, the stability of the banking system--indeed, the stability of the overall **financial** system--is tied critically to the judgments of lenders in the money market.

The most recent experience with a system-threatening run, the collapse of the Continental Illinois Bank in 1984, essentially involved a panic withdrawal by large CD holders. Rumors of insolvency precipitated the run

on the money center bank, which had been funding roughly 90 percent of assets with purchased money (Hetzel 1991). As Greider (1987) describes, the concern of both the Federal Reserve and the FDIC was that, if left unchecked, the Continental crisis could induce a systemwide collapse. Many of Continental's creditors were other banks. More generally, the regulators feared that losses by Continental's creditors might induce runs on a number of other large banks that had been weakened by the 1981-82 recession. It was this fear that induced the banking authorities to intervene in the Continental crisis and protect the uninsured creditors.

* Equity Capital

Finally, Chart 8 illustrates the secular behavior of the ratio of bank equity capital to assets. By definition, bank equity capital equals the difference between assets and liabilities. It specifically equals the sum of common and preferred stock outstanding and undistributed profits. Bank capital is important because it provides a buffer to absorb loan losses. Bank capital/assets ratios must satisfy minimum regulatory standards. From the early 1960s to the early 1980s, the aggregate bank capital/assets ratio dropped by roughly a quarter, from around 8 percent to below 6 percent. The growth in bank assets afforded by the development of the money market (especially over the period 1962-74) was not matched by growth in bank equity.

Since the early 1980s, the aggregate bank capital/assets ratio has climbed, on average. It is important to recognize, however, that this growth is largely a response to increased regulatory pressure in the wake of mounting bank and S&L failures and, relatedly, to new capital standards which have been phased in over the last five years (which we will discuss later). Much of the growth in this ratio also reflects a contraction in the ratio's denominator: assets. Because of the kinds of informational asymmetries stressed by Myers and Majluf (1984), issuing new equity is expensive for banks. Typically, banks use retained earnings to build equity (Baer and McElravey 1993). As a number of studies have indicated (Bemanke and Lown 1991, Peek and Rosengren 1991), meeting capital requirements in recent years has forced many banks to contract asset growth.

Bigger Banks Take Bigger Risks

* Policy Incentives

Another outcome of the collapse of Continental Illinois was that the banking authorities in the United States formally certified the policy of too-big-to-fail. The policy had been implicitly in practice at least since the early 1970s, with the bailout of Franklin National Bank (Hetzel 1991, Boyd and Runkle 1993, and Isaac 1993). However, in September 1984, in the wake of the Continental intervention, the Comptroller of the Currency testified to the U.S. Congress that 11 bank holding companies were too big to fail. Further, in practice, the policy appears to have been extended in varying degrees to banks outside the top 11.(7) As we said earlier, it is important to recognize that the term too-big-to-fail policy actually refers loosely to a menu of policies that vary from lenient treatment at the discount window or in the **valuation** of assets to direct infusion of capital and protection of uninsured creditors.(8)

Plugging one hole in the dike, however, opened up another. The too-big-to-fail policy, of course, indiscriminately subsidized risk-taking by large banks. At the same time, it created a nontechnological incentive for banks to become large. Despite being a well-intentioned effort to protect against liquidity panics in the money market, the policy nonetheless helped create the climate for the 1980s crisis. We will return to this issue repeatedly, since it is fundamental to the policy debate.

* A Size Breakdown

The aggregate bank balance sheets mask some important differences across size classes of banks. Generally speaking, smaller banks adopt more conservative asset and liability positions than do large banks. An important policy issue is whether these differences are due to technological factors or to a regulatory environment that favors large

banks, owing to an operative too-big-to-fail policy. Before we investigate that issue, however, we will document the differences.

Following the convention of the Federal Reserve Bulletin, we divide banks into four size classes, based on the size of their assets: small (those with assets less than $300 million), medium (those with assets between $300 million and $5 billion), large (those with assets greater than $5 billion), and money center (the 10 banks with the largest assets). The data are based on averages over the five-year period 1987-91.

Chart 9 disaggregates loans by these size classes. The share of loans allocated to business lending--the sum of C&I and commercial real estate lending--varies positively with size. (Chart 9 omitted) Both the consumer and residential real estate shares vary negatively. Since business lending generally accounts for the substantial majority of loan losses, the general picture is that larger institutions hold riskier asset positions. Later we will present some information on loan performance that is consistent with this contention.

Chart 10 characterizes the composition of liabilities. (Chart 10 omitted) The key point here is that the relative use of core deposits (checkable and savings and time deposits) shrinks with size, while the relative use of money market instruments increases. About 85 percent of small bank liabilities are core deposits. Conversely, money market instruments constitute roughly 42 percent of large bank liabilities and 54 percent of money center bank liabilities. Further, the money center banks obtain more than half of their purchased funds from abroad. Deposits in foreign offices are mainly money market instruments.) An implication of the differences in liability structure is that larger banks have smaller net interest margins. As Chart 11 illustrates, the net interest margin varies from 3.9 percent for small and medium-sized banks to 2.8 percent for the money centers. (Chart 11 omitted)

In addition to holding riskier asset portfolios and employing greater use of money market instruments, larger banks have lower capital/assets ratios, as Chart 12 shows. (Chart 12 omitted) Indeed, large banks were responsible for much of the secular thinning of the aggregate bank capital/assets ratio portrayed in Chart 8.

Again, a key policy question is, Why? Does this reflect some kind of technological advantage--for example, a better ability to diversify risks or to use scale economies in loan processing? Or does it instead reflect mainly the effect of regulatory bias induced at least in part by the too-big-to-fail policy? We will return to this issue in the next section through an assessment of the recent performance of banks across size classes.

ANALYZING PERFORMANCE

In this section, we describe the bleak performance of banks in the 1980s and pinpoint the reasons for it. We begin by presenting a set of aggregate measures of bank performance. We then turn to an **analysis** of data disaggregated by bank size and location. We determine that even after we controlled for locational effects, performance was significantly related to size. And very large banks did worse than the average.

Bank Profitability Plunges

* Overall

Chart 13 presents the recent trend in two commonly used measures of bank profitability: the rate of return on equity and the rate of return on assets. (Chart 13 omitted) Both measures exhibit similar behavior over the period 1973-91. Both decline fairly steadily after 1979. The exception (for both measures) is a sharp drop in 1987 followed by a sharp recovery in 1988. However, the plunge in 1987 reflects large write-offs of international loans, the timing of which was somewhat arbitrary. The main point of Chart 13 is that after trending down since 1979, bank profitability was, in the latter half of the 1980s, significantly below its average for most of the 1970s. The rate of return on equity dropped from about 14 percent in 1979 to an average of about 8 percent over 1989-91.

Similarly, the rate of return on assets dropped from about 0.75 percent to 0.50 percent.

Chart 14 shows that a rise in the rate of loan losses accompanied the general decline in profitability. (Chart 14 omitted) Provisions for loan losses increased during the 19812 recession, as would normally be the case in a downturn. However, the upward trend in these provisions continued almost throughout the 1980s. The loan loss rate has risen from about 0.2 percent of assets in the late 1970s to nearly 1 percent of assets over the end of the sample period. Conversely, the net interest margin has actually risen slightly over this period, from about 3.3 percent in 1977 to 3.8 percent in the mid-1980s to an average of about 3.5 percent over the last several years. The aggregate measures thus suggest that the decline in bank profitability stemmed from loan losses rather than from a shrinking net interest margin.

Why didn't the net interest margin drop over this period of increased competition and deregulation of interest rates on bank liabilities? In our view, the slight upward trend of the net interest margin is symptomatic of the decline in bank asset quality over the period. That is, the rise in the aggregate loan spread likely reflects the decline in the asset quality mix over the period. The sharp rise in loan losses over the period also fits the general story. In the sections that follow, we will bring more evidence to bear on this issue.

* By Size and Location

It is first useful to provide some background on the cross-sectional distribution of U.S. banks by size and by region of the country.

We divide banks into six asset-size categories based roughly on the classifications used by the FDIC. The asset-size categories range from less than $50 million to more than $10 billion. Chart 15 presents information on the percentage of banks and the percentage of bank assets across the six size classes, based on averages constructed over the period 1983-91. (Chart 15 omitted) Clearly, though there are many thousands of banks in the United States, bank assets are concentrated among a relatively tiny percentage. On average over the nine-year sample, banks with more than $10 billion in assets constituted only 0.3 percent of the total number, but they held 37.4 percent of total bank assets. Banks with more than $1 billion in assets numbered 2.5 percent of the total, but held about two-thirds of the assets. At the other extreme, nearly 80 percent of the banks had less than $100 million in assets, but these banks together only held about 13 percent of the assets.

Chart 16 similarly portrays the distribution of banks across regions used by the U.S. Bureau of the Census. Large banks tend to be located near salt water (on the East and West Coasts), while small banks tend to concentrate around fresh water (in the Midwest). (Chart 16 omitted)

Table 1 presents evidence on U.S. bank failures disaggregated by census region. (Table 1 omitted) Over the period 1980-91, 1,351 banks failed. The peak was the five-year **interval** 1986-90, when nearly 70 percent of the failures occurred. Not surprisingly, there is a strong regional pattern that is closely associated with the temporal pattern of certain regional economic difficulties. The (West South Central) oil states, principally Texas, accounted for the majority of failures: nearly 700. Most of the failures in this region are bunched during 1986-90, roughly the period when oil and real estate prices collapsed there. A distant second in importance are the (West North Central) agricultural states. Agricultural problems in this region during the mid-1980s precipitated over 180 bank failures.

For two related reasons, however, the raw failure numbers portray an incomplete picture. First, these numbers do not take into account the size of failed banks. While small bank failures are far more plentiful, a large bank failure places far greater pressure on the FDIC insurance fund. Despite the rash of failures in the agricultural states, for example, the assets of closed banks never exceeded 1 percent of the total in the region, since virtually all of the banks involved were small. Similarly, despite

there being only 12 bank failures in New England in 1991, assets of failed banks amounted to 8.9 percent of that region's total.

Table 2 confirms the general point. It shows that in the peak period of 1986-90, banks with assets more than $500 million accounted for less than 4 percent of total bank failures, but nearly 60 percent of the total assets of failed banks. (Table 2 omitted) Further, the three banks with assets more than $5 billion that failed accounted for more than 30 percent of the total failed-bank assets.

The second reason the failure data are misleading is that they do not take into account the **historical** regulatory bias in favor of large banks. Because the FDIC has been less willing to close large banks, the failure numbers do not accurately capture overall bank performance.

The biases inherent in using failure data as indicators of bank performance lead us to consider several finer measures. Charts 17 and 18 report, by census region, the ratios of loan loss provisions to assets and net income to assets, respectively. (Charts 17 and 18 omitted) The numbers are averages across individual banks within the individual regions over the period 1983-91. By both indicators, the banks in the West South Central region (which includes Texas) performed worst. Both indicators suggest, however, that regional considerations alone do not provide a complete story. In the troubled New England region, for example, banks on average performed at the national mean in terms of loan loss provisions and only slightly worse than the national mean in terms of net income. Similarly, the poor performance of the money center banks located in the Middle Atlantic region was at variance with other banks there, which performed better than the national mean on average.

Examining the same data by bank asset size yields a clearer picture. Chart 19 reports the ratio of loan loss provisions to assets across the six size classes of banks, and Chart 20 reports the ratio of net income to assets. (Charts 19 and 20 omitted) Across size classes, there is a U-shaped pattern to the loan loss provision ratio. Banks in the largest category (more than $10 billion in assets) performed worst by this measure. The ratio of provisions to assets declines with size, reaching a minimum at the class of banks between $100 million and $250 million in assets. The ratio then begins to rise monotonically as size declines further.

The ratio of net income to assets is, of course, a better overall indicator of performance than is the ratio of loan loss provisions to assets. However, the U-shaped pattern of the latter is simply mirrored by a hump-shaped pattern of the former, as Chart 20 illustrates. Gauged by net income to assets, banks larger than $10 billion still performed worst, and banks between $100 million and $250 million still performed best.

Judging from Charts 19 and 20, we see that relative loan loss performance influences the pattern of net income to assets across size classes. Losses on loans to less-developed countries (LDC loans) were likely an important factor since these losses were heavily concentrated among large banks. Perhaps less well known is that large banks also suffered disproportionately from commercial real estate lending. Recall (from Chart 9) that the fraction of commercial real estate loans in bank portfolios varied positively with size. However, even within the category of commercial real estate lending, large banks performed less well. Table 3 shows that in the third quarter of 1992, the percentage of noncurrant, or what are commonly called problem, real estate loans ranged from about 1.6 percent for the smallest banks to about 7 percent for the largest banks. (Table 3 omitted) For every bank size, business-related real estate lending--the construction and commercial categories--accounted for most of the noncurrent loans, but the share of noncurrent loans within each loan category rises steeply with bank size. Astonishingly, nearly 22 percent of construction loans at the largest banks we noncurrent.

* What's the Problem?

From a policy perspective, it is important to determine how well the negative correlation between size and performance survives, once we control for region. For example, there is a regional dimension as well. as a size

dimension to commercial real estate problems. Table 4 shows that noncurrent real estate loans are heavily concentrated in the Northeast and the West, the two main areas where real estate problems linger. (Table 4 omitted) Thus, it is possible that the correlation between size and performance of the real estate loans is simply due to the fact that the banks in the troubled Northeast and West are larger on average. If this is the case, then restrictions on interstate banking might be primarily responsible for the disproportionate concentration of loan losses. If the negative correlation between size and performance remains after we have controlled for location, then it is possible that distortions induced by a too-big-to-fail safety net may have been important.

* Methodology and the Model

We now investigate the relation between size and performance, after controlling for the influence of region. The data set we use has annual observations on individual U.S. banks over the period 1983-91.(9) The sample is obtained from bank reports of condition and income that are filed with regulators; the sample contains the universe of domestic insured commercial banks over this period.

We consider two performance measures: the ratios of net loan charge-offs to assets and net income to assets.(10) Each, of course, is a measure of ex post performance. Our working hypothesis, particularly for interpreting the behavior of charge-offs, is that over the sample period a poor ex post performance is the consequence of a high degree of ex ante risk-taking. Ideally, we would like to measure ex ante **portfolio** risk. However, this is extremely difficult to do for banks, for two main reasons. First, the sample period is relatively short. Second, the data are based on accounting rather than market **value** measures, and there is considerable evidence that the accounting data are intentionally smoothed. (See, for example, Boyd and Runkle 1993.) This has the effect of causing accounting measures to systematically understate risk. The idea is that during the 1980s there was a series of large negative shocks to the banking system (as Charts 13 and 14 suggest); by examining ex post returns, therefore, we can get some feel for the outer tails of the distributions.

For each bank, we average each of the two performance indicators over the sample period. We work with the time-averaged values of these indicators for several reasons. Firs the timing of charge-offs and income is to some degree arbitrary in the short run. Because clean market **value** assessments of banks' overall **portfolio** are unavailable, banks have some short-run discretion over when they report gains and losses. Over time, the discrepancy between accounting and market **value** indicators declines. Second, working with time-averaged data permits a more parsimonious representation of a model. In general, bank performance should vary over time with economic conditions. However, since we are mainly interested in uncovering secular relationships, it seems reasonable to average out the time effects: the benefit is a much simpler model to evaluate.

At least two types of bias are possible in this study. First, some banks drop out of the sample over time. Since exit is most often due to failure, exit and performance are likely correlated. Omitting exiting banks from the sample could, therefore, bias the estimates. We adjust for this problem by averaging each of the performance indicators for a bank over its lifetime in the sample, even if the bank exits partway through the sample period. In this simple way, we include information from the exiting banks in our estimate

The second type of possible bias is that bank performance could feed back and affect size. If a bank does not perform well, for example, it may decide to contract its assets. We address this issue by using presample data to sort banks. Thus, we use the 1983 data to sort banks into size groups as well as into regions. The performance indicators we use as dependent variables are then time-averaged over 1984-91. For robustness, we also split the sample and work with time averages of the performance variables over 1987-91. In this latter case, we use 1986 data to sort the

banks.

The initial set of independent variables are dummies for census region and for size. We use the FDIC's six size classes defined earlier (in Chart 15). For two reasons, we use discrete indicators rather than a continuous variable for size. First, the earlier descriptive **analysis** suggests that the relationship between size and performance is likely to be highly nonlinear. Second, by using size class indicators which correspond closely to the categories the FDIC uses to report all types of bank data, we directly link our results with a variety of other types of information on bank performance. A similar consideration motivates the use of census regions to denote location: the FDIC uses this indicator to present information on performance across locations.

Here, then, is our formal model. Let (Figure omitted) denote a dummy for region j, (Equation omitted) size dummy for size class k, and (Equation omitted) the time-averaged **value** of a bank performance indicator. Then the basic model we estimate is given by

(1) (Equation omitted)

where (Equation omitted) is a **random** error term and where, to identify the model, one of the coefficients on the six size class dummies is normalized at zero. We model bank performance as a linear function of a region-specific intercept a sub j and a slope coefficient b sub k that depends on the size class of the bank. Under the initial formalization given by equation (1), we restrict the slope coefficient on size class (b sub k) to be identical across regions. We also consider a more general formulation that permits the size class slope coefficients to vary across regions, as given by

(2) (Equation omitted)

Here the slope coefficient on size (b sub jk) is region-specific.

Under the null hypothesis that size is unimportant to performance, the slope coefficients on size equal zero for every size class. If the null is true, then restrictions on interstate banking may be paramount in explaining bank performance. Roughly speaking, if the regional dummies capture all the explanatory power, then it is likely that constraints on the banks' ability to diversify nationally have inhibited banking. But if the too-big-to-fail policy has been a significant distortion (in the context of significant competitive pressures on banking and managerial entrenchment problems for large banks), then we should expect to reject the null. Further, after controlling for regional effects, we should expect an inverse connection between size and performance, especially at the upper tail of the size distribution.

* The Result: Size Matters...

Table 5 reports the results from estimating the basic model, described by equation (1). (Table 5 omitted) There are four regressions, corresponding to two different dependent variables (the ratios of net charge-offs to assets and net income to assets) and two different sample periods (1984-91 and 1987-91).(11) We normalize at zero the coefficient on the banks in size class 3 ($100 million-$250 million in assets). In each case, we easily reject the null that size class is unimportant. Further, to a first approximation, both the U-shaped pattern of loan losses and the hump-shaped pattern of net income across size classes that appeared in Charts 19 and 20 remain after we control for the influence of region.

For the ratio of net charge-offs to assets, the coefficients on the size dummies increase monotonically, moving from size class 3 up to size class 6 (more than $10 billion in assets).(12) Further, this ordering of coefficients is statistically significant, as Table 6 indicates. (Table 6 omitted) An analogous set of results arises when the dependent variable is instead the ratio of net income to assets.(13)

It is also interesting to observe that the smallest banks (those in class 1, with less than $50 million in assets) performed worse than those in the next two larger classes. The difference, however, is sharper on average with the net income ratio than with the net charge-offs ratio. One interpretation is that the smallest banks do not exploit scale economies

that seem available at least up to the class 3 category.

We next turn to the more general model described by equation (2), which permits the slope coefficient on size to vary across regions. Table 7 reports the coefficients on each size class averaged across regions, with the averages weighted by the percentage of banks in the size class of interest that are in the region. (Table 7 omitted) The table also reports the joint significance of a size class dummy across regions for each size class. The results from this general model correspond to those from the restricted one. Once again, both the U-shaped pattern for net charge-offs and the hump-shaped pattern for net income emerge, and both are highly significant. Analogously to Table 6, Table 8 presents tests of the equality of coefficients on adjacent size classes within a region, jointly across all regions. (Table 8 omitted) The message of Table 6 is preserved: between size classes 3 and 6, the inverse ordering between size and performance is significant, and the smallest banks perform poorly relative to those in the two next-larger size classes.

A question that remains is whether the abnormal risk-taking by large banks could be explained by factors completely unrelated to regulatory policy (that is, the subsidy inherent in the too-big-to-fail policy). Could it be the case that for technological reasons large banks have simply specialized in different types of loans than smaller banks and that the large banks have just been unlucky?

We are skeptical of this hypothesis providing a complete explanation, for a variety of reasons. The largest category of banks (those with assets more than $10 billion) performed significantly worse than the next-largest (those with assets from $1 billion to $10 billion). It is hard to believe that important differences in scale economies exist between these two sizes of banks that permit the former to make loans the latter cannot. In addition, the banks in the next size class down, from $250 million to $1 billion, are still reasonably large and thus still relatively unrestricted in the types of loans they can make. In fact, banks in this size category participated in LDC loan syndications. However, they did not typically adopt the same degree of risk exposure as did the larger banks. Indeed, Dombusch (1986) observes that some money center banks held LDC loans equal to twice their capital. More generally, scale economies may explain why only large banks can originate certain types of loans such as LDC loans. However, since loan sales are possible, scale economies do not explain why large banks hold a larger share of these assets on their balance sheets.

A purely technological story also has difficulty explaining why the large banks adopted a riskier liability structure as well as a riskier asset structure. As we documented earlier (in Charts 11 and 12), the large banks operated with both thin equity capital-to-assets ratios and thin net interest margins, in the latter instance due to the extensive use of purchased money. It is worth emphasizing that large bank capital/assets ratios were not only lower than the industry mean, but were also substantially lower than those of competing nonbank intermediaries such as **finance** and life insurance companies (Boyd and Rolnick 1989). A natural explanation for this relative position is that the policy of too-big-to-fail led to a mispricing of the (technically) uninsured liabilities of these institutions.(14)

* ... A Lot

We next conduct a simple experiment to determine the quantitative importance of the poor relative performance of large banks. We compute the reduction in total loan losses that would have resulted if the two largest categories of banks (classes 5 and 6) had performed as well as the third-largest category (class 4). Specifically, for each year and each region, we compute values of net charge-offs for the class 5 and 6 banks, assuming that they had the same net charge-off/assets ratio as the class 4 banks in the same region.(15) We then use this information to compute the yearly reduction in aggregate charge-offs that would have resulted. If the extranormal loan losses of the class 5 and 6 banks reflect the consequences of excessive risk-taking encouraged by regulatory policy, then this

computation is a rough estimate of the cost of this policy.

Table 9 shows that under these assumptions, total charge-offs would have averaged about 25 percent lower over 1983-91. (Table 9 omitted) This amounts to an extra loss in wealth over the period of about $45 billion--if not quite an Okun gap, then certainly a heap of Harberger triangles. To place the number in context, the total equity capital of the banking system is $232 billion. (Charge-offs ultimately reduce capital.) Note that the class 6 banks (those with assets over $10 billion) account for most of the cost. Finally, we observe that two-thirds of the cost--about $30 billion--arises in the peak period of banking difficulties, 1987-91, mainly due to the poor performance of the class 6 banks.

Our cost estimate is conservative, we think, for two reasons. First, we do not use the best performing banks, those in class 3, as the benchmark for calculating the cost. Using the banks in size class 3 ($100 million--$250 million) as the benchmark for performance instead of those in size class 4 ($250 million--billion) would produce a larger estimate.

Second, to the extent that loan losses forced capital constraints to bind tighter, the shadow **value** of charge-offs may exceed the dollar amount. Table 10 presents information by size class on the share of assets held by banks that we capital-constrained during the height of what some call the capital crunch, in 1990 and 1991. (Table 10 omitted) The table shows that the capital crunch was almost exclusively a large bank problem. The banks that were constrained were mainly large, and large banks accounted for nearly all of the assets held by constrained banks. These facts correspond to the recent empirical evidence on the impact of bank capital on loan growth during 1990 and 1991. Both Furlong (1991) and Peek and Rosengren (1991) show that the link between capital declines and loan growth (first documented by Bemanke and Lown 1991) was stronger in magnitude for large banks than for small banks. Thus, to the degree loan losses forced a reduction in lending (via the impact on bank equity), our cost estimate should be adjusted upward. (See also Lown and Peristiani 1993.)

Our calculations are only intended to question the efficiency of the safety net that existed in the 1980s and not that safety net's desirability. As discussed earlier, despite the changes in this industry, a major banking crisis could still potentially disrupt the economy. As Summers (1991) has observed, a **financial** crisis which raised the unemployment rate by one percentage point for one year would result in a $100 billion loss in output.

IN CONCLUSION

In assessing the impact of the too-big-to-fail policy, we find it useful to make the distinction, common in business cycle **analysis**, between impulses and propagation. It is not correct to think of the policy as a primitive causal force, or impulse, in the recent banking crisis. Clearly, the impulses were a series of negative shocks that included defaults on LDC debt and collapsing oil and real estate prices. The too-big-to-fail policy contributed by subsidizing risk-taking and thereby increasing the vulnerability of the banking system to these disturbances.(16) In this way the policy shaped the propagation by creating an environment that enhanced the impact of these impulses. If we accept large banks' extranormal losses as a rough estimate of the impact of this policy, then the cost during the 1980s was $45 billion--or about 20 percent of the capital of the banking industry. An additional cost (even harder to quantify) was that large banks were the main culprits in the 1990-91 capital crunch.

Making the distinction between impulses and propagation clarifies the fact that it is not meaningful to simply argue that bad luck was responsible for the plight of the large banks. Of course, it is the case that large banks were unlucky, since they were heavily invested in assets which experienced negative shocks during the 1980s. However, a similar statement could be made about the savings and loans. They were unlucky in an analogous way. With large banks as with the savings and loans, the key

issue is whether the **portfolio** structure these **financial** firms adopted was distorted by regulatory bias. As we have discussed, it is hard to believe that the **portfolio** structure of very large banks (for example, heavy investment in LDC and commercial real estate lending, in conjunction with thin capital/assets ratios) could be explained simply by scale economies. To us this becomes particularly apparent when one examines the behavior of banks above $10 billion in assets, a range in which scale economies are no longer likely to exist.

What are the implications of our **analysis** for the recently introduced policy reforms? The most significant reforms, of course, are the Basle Accord of 1988, which introduced risk-based capital standards, and the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA). Both reforms appear to directly confront what our **analysis** suggests has been the main problem: the subsidy to risk-taking by large banks. As we have argued, an important way the subsidy has played out has been that large banks have held less capital than they might have otherwise. The new Bank for International Settlements capital requirements should help offset this distortion and, in this way, force the large banks to better internalize the costs of their **portfolio** decisions. And the increased cushion of capital reduces the probability that taxpayers will have to **finance** loan losses.

Provisions of the FDICIA attempt to roll back the too-big-to-fail policy. Saving a large U.S. bank now requires the formal concurrence of bank regulators, the secretary of the Treasury, and even the president. These provisions also restrict discount window lending, a favorite tool used over the last decade to keep troubled large banks afloat. Finally, and importantly, the provisions impose restrictions on interbank lending to banks that fail to meet adequate capital standards. The goal here is to reduce the likelihood that closing a large bank will precipitate a wave of failures throughout the banking system. The idea is to avoid the kind of trap regulators fell into during the Continental Illinois crisis.

Whereas we are generally optimistic, the new policy regime has not really been tested yet, and it is impossible to predict how it will ultimately work. Moreover, the new regime contains a number of provisions which are undesirable in our view. However, these issues are beyond the scope of the present study.

Finally, we discuss some of the implications of our **analysis** for two related policy issues--namely, interstate branching and bank mergers. Our tests indicate that very small banks also did rather poorly during the 1980s. This finding suggests that the inability to exploit scale economies, rather than disproportionate loan losses, may be the main problem for the smallest category of banks (those with less than $50 million in assets). Encouraging these banks to merge with larger banks may therefore be desirable. At the same time, we are skeptical about the benefits of permitting mergers among very large banks. The clear pattern of our results is that in the 1980s banks in the middle of the size distribution (with assets of $100 million--$1 billion) performed best. Several detailed studies of the issue have also concluded that recent large bank mergers have not produced efficiency gains (Berger and Humphrey 1991 and references therein).

Restrictions on interstate banking have likely contributed to the high number of bank failures, particularly failures of small banks in the oil and agricultural regions. But as we have argued, the main stress on the system has not been the raw number of bank failures; rather, it has been the poor performance of large banks. Restrictions on interstate banking do not prevent large banks from diversifying their loan portfolios nationally. Specifically, these restrictions do not preclude banks from opening loan production offices across state borders. While scale economies may inhibit smaller banks from pursuing this activity, large banks do not face formidable obstacles to national (or even international) lending. Accordingly, we do not think that interstate branching restrictions have

been primarily responsible for bank failures. We do think, however, that there is a case for further reducing restrictions on interstate banking. It is true that branching facilitates lending to smaller borrowers. In this vein, branches may be more efficient conduits than loan production offices for cross-state lending. Any reform which improves the efficiency of large banks is worth taking seriously.

FOOTNOTES

(1) We emphasize that the positive correlation between access to the large certificate of deposit (CD) market and size is not clear evidence of a technological factor that justifies large banks holding low capital. This is because the access may be aided by regulatory policy which implicitly guarantees large banks' money market liabilities. After all, the Continental Illinois Bank enjoyed a period of great access to the money markets. as we discuss below.

(2) Another very important development in banking over this period has been the movement of bank activities off-balance sheet. This development is considered in more detail in Boyd and Gertler 1993.

(3) Roughly speaking, we can divide commercial loans into two categories: those made to smaller, less well-known firms, which require evaluation and monitoring, and those made to highly rated firms, which require relatively little information-processing. The former are typically priced off the prime lending rate, while the latter are typically priced off the cost of issuing large CDs, the banks' marginal source of funds.

(4) Underlying the growth of commercial real estate lending were both tax incentives and relaxation of regulatory constraints on banks in the early 1980s. Subsequent reversals of the tax incentives contributed to the decline in real estate. For details, see Hester 1992 and Litan 1992.

(5) Some qualification is in order since government-sponsored securitized mortgages are treated as securities rather than mortgages in intermediary accounting statements. We thank Myron Kwast for pointing this out.

(6) Checkable deposits include demand deposits and retail transaction deposits such as negotiable order of withdrawal (NOW) accounts. While NOW accounts pay interest, their rates appear much less sensitive to market forces than rates on other interest-bearing bank liabilities.

(7) O'Hara and Shaw (1990) present evidence that news of the Continental bailout policy raised the stock prices of large banks, but not the stock prices of small banks [which O'Hara and Shaw (1990, p. 1588) dubbed too-small-to-save].

(8) We are not suggesting that the too-big-to-fail policy completely eliminated market discipline over large banks. Indeed, the management of Continental Illinois was fired. One should not focus on this policy in isolation. As we will discuss later, we believe the combined climate of too-big-too-fail, competitive pressures on banking, and possibly problems of managerial entrenchment (as discussed in Boyd and Graham 1991 and Gorton and Rosen 1992) contributed to the substantial rise in risk-taking by large banks.

(9) The organizational entities we study here are banks. It is true that many banks are owned by bank holding companies, which control one or more banks and often non-bank affiliates as well. For many purposes, the most appropriate organizational entity is the consolidated holding company. However, the objective here is specifically to study bank performance. Consolidated statements for holding company banks are not easily available. Finally, though they do not control for regional effects. Boyd and Runkle (1993) do obtain evidence of an inverse relation between performance and holding company size that is similar to the inverse relation between performance and size which we find at the bank level.

(10) Net loan charge-offs include all loans determined to be uncollectible net of recoveries on (previously written-off) loans. This entry is not an accounting expense, but rather a reduction in a reverse account. Provision for loan losses is the accounting expense entry which reduces profits. When the data are averaged over several years, as they are

here, the two loan loss measures are highly correlated. Thus, for present purposes, it makes little difference which is employed.

(11) Though we do not report the statistics here, the general results we obtain are robust to using the first half of the sample period, 1984–86, and also to running the regressions year by year.

(12) the results are the same if as the dependent variable we use net charge-offs divided by loans rather than net charge-offs divided by assets. We chose the latter because we are interested in analyzing the ex post performance of the entire bank **portfolio**.

(13) Because equity is measured in book values, we do not consider the rate of return on equity as an alternative dependent variable. Since this measure does not include capital gains and losses on equity, it could be seriously distorted. For example, a bank with near zero equity due to poor performance could have a high ratio of net income to equity.

It is true that there is a size bias in the ratio of net income to assets, since large banks use systematically more **financial** leverage. However, a reasonable calculation suggests that this bias is small relative to the differences we observe in the data.

(14) In Boyd and Gertler 1993, we include some empirical results which, due to space limitations, are not reproduced here. In particular, we show that the size effect remains significant after we include **portfolio** share variables--for example, the ratio of commercial and industrial loans to total loans and the ratio or commercial real estate loans to total loans. However, it is difficult to interpret these regressions. Suppose, for example, that too-big-to-fail induced a large bank to invest more heavily in commercial real estate lending. Because the loan share variable is continuous, it might do better in explaining loan losses than the size dummy which was the true primitive factor. Hence, even if we had found that loan shares displaced size, it would not necessarily be evidence against our hypothesis.

(15) We are assuming that class 4 bank portfolios are available in elastic supply (that is, that the type of **portfolio** held by class 4 banks is available in elastic supply to class 5 and 6 banks).

(16) This statement presumes that it is possible to have a banking safety net in the absence of the particular too-big-to-fail policy that prevailed over the last decade. We expand on this issue a few paragraphs below.

REFERENCES

Baer, Herbert L., and McElravey, John N. 1993. Capital shocks and bank growth--1973 to 1991. Economic Perspectives 17 (July-August): 2–21.

Berger, Allen N., and Humphrey, David B. 1991. The dominance of inefficiencies over scale and product mix economies in banking. Journal of Monetary Economics 28 (August): 117–48.

Bemanke, Ben S., and Lown, Cara S. 1991. The credit crunch. Brookings Papers on Economic Activity 2:205–39.

Boyd, John H., and Gertler, Mark 1993. U.S. commercial banking: Trends, cycles, and policy. NBER Macroeconomics Annual 1993: 319–68.

Boyd, John H., and Graham, Stanley L. 1991. Investigating the banking consolidation trend. Federal Reserve Bank of Minneapolis Quarterly Review 15 (Spring): 3–15.

Boyd, John H., and Rolnick, Arthur J. 1989. A case for reforming federal deposit insurance. Federal Reserve Bank of Minneapolis 1988 Annual Report: 3–16.

Boyd John H., and Runkle, David E. 1993. Size and performance of banking firms: Testing the predictions of theory. Journal of Monetary Economics 31 (February): 47–67.

Dornbusch, Rudiger. 1986. Dealing with debt in the 1980s. International trade and **finance**, ed. Robert Baldwin and J. David Richardson. Boston: Little, Brown.

Federal Deposit Insurance Corporation (FDIC). 1993. 1992 Annual Report.

Furlong, Frederick T. 1991. Can bank capital regulation work? Research

revisited. Federal Reserve Bank of Son Francisco Economic Review (Summer): 32-33.

Gorton, Gary, and Rosen, Richard 1992. Corporate control, **portfolio** choice, and the decline of banking. **Finance** and Economics Discussion Series 215. Board of Governors of the Federal Reserve System.

Greider, William. 1987. Secrets of the temple: How the Federal Reserve runs the country. New York: Simon and Schuster.

Hester, Donald 1992. **Financial** institutions and the collapse of real estate markets. Manuscript. University of Wisconsin.

Hetzel, Robert L. 1991. Too big to fail: Origins, consequences, and outlook. Federal Reserve Bank of Richmond Economic Review 77 (November/December): 3-15.

Isaac, William M. 1993. Saving the banks: Limiting damage of deposit insurance. Wall Street Journal (March 4): A14.

Litan, Robert E. 1992. Banks and real estate: Regulating the unholy alliance. Manuscript. Brookings Institution.

Lown, Cara S., and Peristiani, Stavros. 1993. The 1990 credit slowdown: Did large banks create a credit crunch? Manuscript. Federal Reserve Bank of New York

McCauley, Robert, and Seth Rama. 1992. Foreign bank credit to U.S. corporations: The implications of offshore loans. Federal Reserve Bank of New York Quarterly Review 17 (Spring): 52-65.

Myers, Stewart C., and Majluf, Nicholas S. 1984. Corporate **financing** and investment decisions when firms have information that investors do not have. Journal of **Financial** Economics 13 (June): 187-221.

O'Hara, Maureen, and Shaw, Wayne. 1990. Deposit insurance and wealth effects: The **value** of being "too big to fail." Journal of **Finance** 45 (December): 1587-1600.

Peek, Joe, and Rosengren, Eric. 1991. The capital crunch: Neither a borrower nor a lender be. Working Paper 91-4. Federal Reserve Bank of Boston.

Summers, Lawrence H. 1991. Planning for the next **financial** crisis. In The risk of economic crisis, ed. Martin Feldstein, pp. 135-58. Chicago: University of Chicago Press.

* This article is an abbreviated version of "U.S. Commercial Banking: Trends, Cycles, and Policy," a paper published in the NBE Macroeconomics Annual 1993, ed. Olivier Jean Blanchard and Stanley Fischer, pp. 319-68, Cambridge, Mass.: MIT Press. This article appears here with the permission of the MIT Press. The authors thank, for help, Allen Berger, Fischer Black, Olivier Blanchard, Martin Feldstein, Stan Fischer, Chris Flinn, Jordi Gali, Simon Gilchrist, Stan Graham, Stuart Greenbaum, Jeff Gunther, Ron Johnson, David Jones, Myron Kwast, Cara Lown, Franco Peracchi, Stavros Peristiani, Ken Robinson, Art Rolnick, David Runkle, Gary Stern, and Ken Wolpin. They also thank Gabriele Galati and Joel Krueger for outstanding research assistance.

John H. Boyd** Senior Research Officer Research Department Federal Reserve Bank of Minneapolis

Mark Gertler Professor of Economics New York University and Research Associate National Bureau of Economic Research

** Also, Adjunct Professor of **Finance**, University of Minnesota.

Special Features: Charts; Graphs; Equations; References
Classification: 1110 (CN=Economic conditions & forecasts); 8110 (CN=Commercial banking); 9130 (CN=Experimental/Theoretical); 9190 (CN=United States)

Descriptors: Commercial banks; Deregulation; Competition between **financial** institutions; Bank failures; Size of enterprise; Loan losses; Economic conditions; Studies
Geographic Names: US

36/9/22 (Item 22 from file: 625)
0058965
The Federal Reserve Board series of proposals

American Banker Plus - January 30, 1987 ; Pg. p ; Vol. 151 , No. 21
Article Type: News

Text:

The Federal Reserve Board today issued for comment a series of proposals to reduce and control the payments system risk faced by the Federal Reserve and individual depository institutions participating in large-dollar wire transfer networks, book-entry transfer systems, and automated clearing houses (ACHs). These proposals supplement the payment system risk policy announced by the Board on May 17, 1985.

Large-dollar funds transfer networks are an integral part of the payments and clearing mechanism, and total overdrafts by participants making funds transfers on these networks frequently exceed more than $80 billion daily. A daylight overdraft occurs when an institution sends funds over Fedwire in excess of the balance in its reserve or clearing account, or sends more funds over a private network than it has received.

An other $60 billion or more of overdrafts occur each day from book entry transfers of U.S. Government and agency securities; book-entry overdrafts occur when a depository institution receives (and pays for) more securities over Fedwire than it has sent (and been paid for).

The Board's basic policy is designed to reduce the potential for Federal Reserve losses and systemic risk to the banking system associated with settlement failure through a reduction, over time, in both the total volume of daylight overdrafts and the number of institutions with a pattern of substantial reliance on such credit.

Changes proposed by the Board in its payment system risk policy, with a 60-day comment period, would:

* provide depository institutions incurring daylight overdrafts as a result of book-entry government securities transactions with two options. First, depository institutions could include these overdrafts with net debits that arise from transfers of funds over Fedwire and private wire networks for purposes of determining the total debits subject to the ceiling on the amount of intra-day credit an institution can incur (net debit cap). Alternatively, depository institutions could collateralize their book-entry related overdrafts with eligible book-entry securities that the institutions receive over Fedwire and only the uncollateralized portion would be added to the net debit subject to the cap. Further, the Board is seeking specific comment on whether there should be a book-entry transfer limit and whether that limit should be $25 or $50 million. The Board plans to make the new policy effective on March 23, 1988.

* reduce the levels for the net debit cap established in May 1985 by 25 percent effective June 18, 1987. At the end of 1987 the Board will consider whether to reduce the cap further.

o establish a new de minimis cap category for institutions that do not incur large or frequent daylight overdrafts. This cap would be the lesser of 10 percent of capital or $500,000 and would be available to institutions that do not undergo the self-assessment required for establishing a positive net debit cap under the Board's policy. The

institution's board of directors would have to approve the cap.

o amend its policy statement on payment system risk to indicate that efforts by holding companies to consolidate payment activities at one subsidiary through affiliate transfers over Fedwire that create a pattern of overdrafts at the sending institution would either (a) be prohibited or (b) permitted only under certain conditions consistent with safe and sound banking practices. Comment on the above proposals should be received by February 9.

In addition, the Board proposed several changes affecting automated clearing house (ACH) procedures. These changes include: (1) for purposes of calculating daylight overdraft levels only, posting all entries for ACH debit payments and checks as-of 1 p.m. Eastern time, (2) granting finality for ACH credit payments of $5,000 or less at 1 p.m. local time, and (3) treating as provisional credit all ACH debit and those ACH credit items over $5,000 until the Reserve Banks have actually received the funds. Comment for there changes is requested by March 16.

In connection with its proposals on daylight overdrafts, the Board is seeking comment on the concept of charging a fee for all daylight overdrafts, the Board is seeking comment on the concept of charging a fee for all daylight overdrafts in accounts maintained with the Federal Reserve that are subject to the net debit cap. The objective of this fee would be to provide an additional incentive for depository institutions and their customer to adopt policies and procedures that would reduce daylight overdrafts. Comment is requested by April 13.

Orders detailing the above proposals are attached.

Attachement

FEDERAL RESERVE SYSTEM

(Docket No. R-0587)

Request for Comments on Proposals Regarding

Payment System Risks

Book-Entry Securities Transfers

AGENCY: Board of Governors of the Federal Reserve System. ACTION: Request for comment.

SUMMARY: The Board is proposing for public comment a policy for reducing the risks arising from daylight overdrafts associated with transfers of book-entry securities on Fedwire. The proposed policy includes the following principal components:

1. Depository institutions (and other entities, such as U.S. branches and agencies of foreign banks) would choose between including all book-entry overdrafts with their net debit positions arising from cross system funds transfers for determining their total overdrafts subject to their net debit caps, or collateralizing book-entry overdrafts with the eligible incoming book-entry securities and including only the uncollateralized portion of their book-entry overdrafts with the cross system funds overdrafts subject to their caps.

2. Each institutions choosing the collateralization option would enter into a written security agreement with its Reserve Bank and warrant that a specified minimum percentage of book-entry overdrafts would always be covered by collateral. In monitoring an institution's compliance with the Warranty, a Reserve Bank would apply a margin to the **value** of the securities to account for interest rate and clearing risk. The margin for clearing risk would be established for each institution choosing this option based on a self-evaluation conducted according to Board established guidelines.

3. Book-entry securities transfers on Fedwire (with the exception of original issue transactions and stripped securities) would be subjected to a transaction size limit of either $50 or $25 million, with public comment solicited on the most appropriate level.

DATE: Comments must be received by February 9, 1987. The Board expects that the policy will become effective on March 23, 1988.

ADDRESS: Comments, which should refer to Docket No. R-0587, should

be addressed to the Board of Governors of the Federal Reserve System, 20th and C Streets, N.W., Washington, D.C. 20551, Attention: Mr. William W. Wiles, Secretary; or delivered to Room B-2223 between 8:45 a.m. and 5:15 p.m. Comments received may be inspected in Room B-1122 between 8:45 a.m. and 5:15 p.m., except as provided in S 261.6(a) of the Board's Rules Regarding Availability of Information, 12 C.F.R. S 261.6(a).

FOR FURTHER INFORMATION CONTACT: Edward C. Ettin, Deputy Director (202-452-3368), or Stephen A. Lumpkin, Economist (202-452-2378), Division of Research and Statistics; Elliott C. McEntee, Associate Director (202 452-2231), Division of Federal Reserve Bank Operations; Oliver I. Ireland, Associate General Counsel (202-452-3625), or Joseph R. Alexander, Senior Attorney (202-452-2489), Legal Division; Board of Governors of the Federal Reserve System, Washington, D.C. 20551. For the hearing impaired only: Telecommunications Device for the Deaf (202-452-3544), Earnestine Hill or Dorothea Thompson. SUPPLEMENTARY INFORMATION:

This is one of a series of proposals regarding payment system risk that the Board is issuing for public comment today. The others concern the net debit cap (Docket Nos. R-0588 and R-0589), pricing of daylight overdrafts (Docket No. R-0592), consolidation of affiliated institutions for cap monitoring purposes (Docket No. R-0590), and treatment of payments processed through automated clearing houses (Docket No. R-0591). The Board encourages all interested parties to comment on each of these proposals.

The Board urges that in filing comments on these proposals, commenters prepare separate letters for each proposal, identifying the appropriate docket number on each.

This procedure will facilitate the Board's processing and **analysis** of the comments on these complex proposals, and will ensure that each comment is quickly brought to the attention of those responsible for analyzing the proposal.

BACKGROUND

In May, 1985, the Board announced its policy to reduce the risks that large-dollar payments systems present to the Federal Reserve, to depository institutions and other entities (such as U.S. branches and agencies of foreign banks and Edge Act corporations) using such systems (hereafter referred to as "institutions"), to the banking system, and to other sectors of the economy. 50 Fed. Reg. 21,120 (May 22, 1985). In formulating this policy, the Board was concerned about the effect that overdraft restrictions could have on the U.S. government securities market, the smooth functioning of which is vital both to the conduct of monetary policy through Federal Reserve open market operations and to the efficient funding of the federal debt. Consequently, the Board exempted from quantitative overdraft controls, such as sender net debit caps, Fedwire daylight overdrafts resulting from the transfer of book-entry securities. Rather, the Board sought comment on a proposal to control the risks associated with such overdrafts by requiring institutions incurring them to choose one of three collateralization options. 51 Fed. Reg. 21,132 (May 22, 1985).

Comments on these proposals were largely negative, and the Board's staff reevaluated the proposals. As a result of this reconsideration, together with discussion with industry groups, new collateralization options were developed, supplemented by other proposals not previously considered. These staff recommendations formed the basis for the proposals on which the Board is now seeking comment. Full details on the staff's recommendations, including a detailed comment summary, an **analysis** of the markets in Treasury and agency securities, an evaluation of policy options, and likely market responses, may be found in the staff study, Book-Entry Daylight Overdrafts (Nov. 1986). Copies of this study are available free of charge from the Secretary of the Board at the address noted above, or from the Daylight Overdraft Liason Officer of each Federal Reserve Bank. The Board encourages all parties interested in commenting to

obtain a copy of the staff study, as it contains background information that may enable them more readily to understand the rationale for the Board's proposals.

Because the issues associated with these proposals have been subjected to comment previously and interested parties are familiar with them, the Board believes that a 60-day comment period is sufficient. Further, in order to provide the public with time to prepare for implementation, the Board plans to implement this new policy on March 23, 1988, unless the public comments reveal substantial, unforeseen difficulties with the Board's proposal that require a significantly different policy.

PROPOSAL

Introduction

The Board's May, 1985, proposals would have required institutions incurring book-entry overdrafts to select one of three collateralization options: (1) Treat book-entry overdrafts the same as other daylight overdrafts, subjecting both to the sender net debit cap; (2) establish a stable pool of collateral to secure book-entry overdrafts; or (3) establish a pledge account containing securities, including customer securities, that could be pledged to collateralize book-entry overdrafts. Under the third option, institutions would have been required to shift securities out of the pledge account when they no longer became eligible to pledge, say, as a result of a payment for the securities by an institution's customer.

The proposal on which the Board is now requesting comment would continue to permit institutions to choose the first option; the other two options, however, have been dropped. In place of the two collateralization options, the Board is proposing a modified pledge account option that will allow an institution to collateralize book-entry overdrafts with the incoming book-entry securities (other than paid for securities and securities not eligible to pledge) and to include with cross-system funds overdrafts subject to the sender net debit cap only that portion of the book-entry daylight overdraft that is not so collateralized. For ex post monitoring purposes, each institution choosing the collateralization option would warrant to its Federal Reserve Bank that a specific minimum percentage of its book-entry overdrafts would be collateralized by securities in the pledge account. A margin would be applied to the **value** of the securities offered as collateral to represent risks to the Reserve Banks of (1) declines in collateral values and (2) deficiencies in the pledgor institution's internal operating controls over its securities transfer and clearing business. The Board's policy is rounded out by (1) the use of a supplementary self-assessment of each institution's own book entry operations and controls as a factor in determining margin amounts, and (2) a maximum limitation on Fedwire book-entry transactions of either $25 or $50 million.

Details of the Board's proposal follow:

1. Collateralization

A. Pledge Agreement

A depository institution or other entity choosing to collateralize its book-entry overdrafts would take two steps. First it would enter into a written agreement with its Reserve Bank granting the Reserve Bank a security interest in all those securities that are eligible collateral under the Board's policy. A Reserve Bank's actual collateral position at any time during the day would be determined by the total market **value** of the book-entry securities eligible to be pledged under the Board's policy. Given accounting lags, market price changes, and possible conflicting interests in the securities, a Reserve Bank would know the actual **value** of the securities in which it could successfully assert a security interest only after the fact.

Second, recognizing the impossibility of tracking the exact collateral amount that secures a book-entry overdraft at any point in time, the institution would warrant to its Reserve Bank that the adjusted **value**-- of its pledgeable securities would be no less than a stated
warranty ratio.-- This ratio would be used to determine collateralized and uncollateralized book-entry overdraft amounts for day-to-day ex post cap monitoring purposes. As described in detail below, the relationship of the warranty to actual collateral values would be checked on a periodic basis.

B. Eligible Collateral

The Board believes that institutions should be permitted to pledge only securities that were received through book-entry transfers to secure book-entry related overdrafts. A healthy institution should not be permitted to pledge **portfolio** assets or securities released each day from
pledge as collateral for dealer loans and maturing hold-in-custody and three-party repurchase agreements (RPs). Accordingly, the Board proposes to permit institutions to count as collateral for book-entry related overdrafts only those book-entry securities that the institution is authorized to pledge and that are transferred to the pledging institution over Fedwire on the particular day they are pledged.

The Board realizes that excluding collateral released from maturing RPs and loans may require an increase in costs to depository institutions to track throughout the day those securities in a dealer's position that do not come in during the day on the book-entry wire. Therefore, the Board is requesting comment on what the increase in costs for institutions such tracking is likely to be. Specifically, the Board is interested in knowing the cost to clearing banks of excluding from their own customers' collateral used to secure credit extensions those securities that were not transferred over the book-entry wire that day--and hence did not give rise to a book-entry overdraft.

C. **Value** Adjustments

In order to protect Reserve Banks against credit exposures and to increase incentives for institutions to improve prudential controls over (and reduce the size of) their book-entry overdrafts, the Board proposes two adjustments to the **value** of eligible-to-pledge collateral.

First, for purposes of book-entry collateralization only, a market risk adjustment would be subtracted to protect the Reserve Banks against interest rate changes over the **interval** between the time the collateral is
taken and the time the Reserve Bank's claim is extinguished. The Board is proposing to adopt a market risk "haircut' of between three and five per cent to be applied to book-entry securities collateral on a daily basis. Reserve Banks would be given the flexibility to choose haircut factors within this range for purposes of applying the standard to particular institutions. The size of the individual institution market-risk haircut could be reviewed as often as the Reserve Bank wishes, but on any day it would be fixed within the three to five per cent range. Given recent price history, a daily haircut for market risk in this range should be sufficient to account for most day-to-day fluctuations in prices of government securities. The Board specifically requests comment on whether this procedure is desirable, whether a fixed haircut should be applied to all institutions, or whether the criteria for application of the haircut to institutions should be further refined.

A supplementary haircut in addition to the market risk margin, which would be specific to each institution's own operations, would be based on the results of each institution's self-assessment of these risks. This haircut, which would be subject to supervisory review, is initially expected to be between 0 and 10 per cent for institutions with excellent to satisfactory assessments.

The self-assessment of an individual institution's controls and procedures in its book-entry operations would be an extension of the self

assessment approach of the earlier Policy Statement, which addressed policies, procedures, internal controls, and monitoring capabilities. Under the proposed policy on book-entry risks, four basic areas would be addressed in detail as they relate to book-entry clearing and settlement activities:

- o Credit policy and controls;
- o Collateral monitoring and control;
- o Operational risk; and
- o Funding capacity.

The self-assessment of credit policy and controls would look in detail at the adequacy of an institution's policies and procedures for establishing credit limits for a customer or a group of related customers and monitoring the intra-day exposures within these limits. Although the focus of the monitoring would be heavily on the book-entry activity, the institution's overall exposure to the customer would also be taken into account.

The assessment of an institution's ability to monitor the position of a customer's collateral would focus on this critical element of the institution's exposure in book-entry activity. A sound credit judgment would be impossible without both a good measure of control over what collateral is available to secure a customer's overdraft position in book entry securities.

The assessment of an institution's operational environment would have to identify risks posed by such factors as capacity constraints, internal bottlenecks, and other operating conditions that (1) could affect internal information flows needed to make otherwise sound policies and procedures work properly, or (2) could affect the overall operation of the book-entry securities market and the exposures of the institution itself and other institutions in the market. The reliability of automated systems, the availability of back-up processing capability, and the ability to reconcile and resolve fails and suspense items would be key factors in this area.

Finally, the assessment of funding capacity would look at the ability of the institution to tap the funds market to support not only its normal level of funding needs, but also its ability to fund large book-entry securities positions of its own or its customers in situations involving temporary operational disruptions or external market strains. In this regard, market perception of the institution, existing and normal funding patterns, demonstrated funding capacity, and identified contingency funding plans are key factors.

Each of the four factors (credit policy and controls, collateral monitoring and control, operational risk, and funding capacity) would be rated on a four level scale of Excellent, Very Good, Satisfactory, and Unsatisfactory, with an overall summary rating. Any institution rated unsatisfactory on any of the four factors would not be eligible to participate in the option permitting collateralization of overdrafts related to book-entry activity, and all of its book-entry overdrafts would be included with cross-system funds overdrafts for purposes of the consolidated net debit cap.4/ Institutions with an overall rating of excellent would require no additional haircut on their pool of eligible collateral after the adjustment for market risk. Institutions rated satisfactory on all four factors would take an additional haircut of 10 per cent. Those with a very good rating would take an additional haircut of 5 per cent.

Further details on this self-assessment procedure may be found in the staff study referred to earlier.

The Board requests comments on whether the additional haircut to cover these risks is needed, and whether the self-assessment guidelines that the Board is proposing are appropriate.

D. Warranty

As part of the collateralization agreement, the pledging institution would warrant to its Reserve Bank that a specific percentage of its book

entry related overdrafts would always be covered by eligible collateral as adjusted. This warranty ratio would be used for cap monitoring purposes only, i.e. for determining the uncollateralized daylight overdraft that would be subject to the cap. The Board estimates that banks providing clearing services for broker-dealers should have adjusted-collateral-**value**
to-book-entry-overdraft ratios of 85 to 95 per cent, and thus would be able to use warranty ratios of at least that amount.

The warranty ratio selected by each institution would be based on the **historical** evidence of the adjusted values of eligible-to-pledge securities relative to its book-entry overdrafts. Each institution would have to present evidence to its Reserve Bank to support or modify its warranty ratio; the Reserve Bank would be able to change that ratio if the Reserve Bank's independent review called for it. This review could take several forms. The normal periodic examination would, for example, test the warranty and review the margin for other Reserve Bank risks associated with the self-evaluation guidelines. Moreover, on a **random**
basis--say
twice a month for clearing banks--the Reserve Bank would ask the institution at the end of the day to demonstrate ex post that the adjusted **value** of its eligible collateral in its accounting record at a
specific
time that day was equal to or larger than the warranty percentage of its book-entry overdrafts. If it was not, the Reserve Bank might lower the warranty ratio, pending new evidence from the institution. Thus, spot checks, as well as periodic certification coupled with normal examination, would provide checks on the adequacy of the warranty.

With the warranty ratio used only for monitoring purposes and the real collateral position coming from the repledging of eligible incoming securities, there would be no need to require institutions to reposition collateral between accounts at Reserve Banks, as under the pledge account option published for comment in 1985.

Under the proposal, each institution would be given the choice--but not be permitted to switch back and forth-- of either (1) using the warranty percentage throughout each day, or (2) adjusting the intra-day warranty amount the next day by providing its Reserve Bank with the measured adjusted dollar amount of pledgeable collateral the institution held each 15 minutes during the day. Institutions choosing the second option could have the benefit of eligible collateral in excess of their own warranty when they could demonstrate it. They would also bear the cost -higher uncollateralized overdrafts subject to cap--when the collateral data available the next day indicated a level below their warranty ratio. This approach would provide an incentive for institutions to develop collateral tracking programs in order to be able to show collateral positions above their minimum warranty ratio and thus lower their overdraft subject to cap. At each institution, the collateral tracking data would only have to be recaptured ex post. Under either approach, Reserve Banks would have to compare from time-to-time warranty ratios (or amounts) with the ex post adjusted **value** of pledgeable
securities for
which a security interest has been taken. The actual pledged securities would be the same under both approaches.

E. Consolidated Net Debit Cap

The voluntary sender net debit cap--now applicable to cross-system funds overdrafts--would, under the proposed policy, become a cross-system consolidated net debit cap applicable to the sum of cross-system funds and uncollateralized book-entry overdrafts. Institutions would continue to establish their own caps through a self-evaluation based on current Board guidelines; those institutions not adhering to the guidelines and the policy would, as under the present policy, be prohibited from incurring funds transfer overdrafts on Fedwire. Under the new policy, they would also not be able to incur book-entry overdrafts. The Board would also

consider it an inappropriate use of Fedwire to substitute purposefully book-entry transfers (which can be collateralized) for funds transfers (which cannot) in order to avoid the constraints of the consolidated sender net debit cap.

The current daylight overdraft policy authorizes a Reserve Bank to take full collateral for Fedwire funds overdrafts whenever it believes it is necessary to protect its own position with an individual institution. The Board proposes that this policy be extended for book-entry overdrafts at problem institutions as well, permitting Reserve Banks to take other collateral for book-entry overdrafts, if deemed necessary. Full collateralization is required by the Board's current policy for all the Fedwire funds overdrafts of Edge corporations, bankers' banks, institutions with negative adjusted primary capital, and for the amount by which the Fedwire funds overdrafts of U.S. branches and agencies of foreign banks exceeds their cap based on their "U.S. capital equivalency." The Board proposes that the required collateralization for Fedwire funds transfers for these special entities be extended to book-entry overdrafts as well.

2. Transfer Limits

The Board is also proposing to adopt a mandatory size limit on book entry securities transfers of either $25 or $50 million. The Board believes that this limit would not change market trade size, but would likely alter delivery practices so that transactions would be split, and partial delivery of orders could begin earlier in the day. The Board estimates that this size limit would increase transactions by less than 10 per cent of all transactions, but would affect about one-third of the dollar **value** of book-entry transfers. The objective of the transfer limit
is to constrain intra-day position-building by dealers, spread book-entry volume more evenly over the day, and limit the level of book-entry overdrafts. Maximum transfer limits would not apply to either original issue transactions or to transfers of stripped securities.

These limitations would only be effective if sellers and purchasers of securities are willing to accept and pay for multiple transactions, know what their rights are in the case of a failure to deliver one or more transactions involved in a single trade, and do not simply continue to build positions thereby increasing the size and duration of overdrafts and contributing to end-of-day volume bottlenecks. Thus, to ensure that transfer size limits are effective, the staff of the Board and the Federal Reserve Bank of New York will work with the various committees of the Public Securities Association, as well as other industry representatives, to encourage development of conforming delivery practices and compensation rules.

The Board is requesting public comment on whether the proposed maximum transaction limit should be $25 or $50 million. With a lower limit, position building may be minimized, but such a limit may unduly increase the transactions costs of large trades and have negative market effects.

If the Board finally adopts this policy of a size limit of either amount, it will modify its current policy on proper uses of Fedwire. On March 29, 1984, the Board issued a policy statement stating that "use of Fedwire for the avoidance of Federal Reserve or private sector risk reduction measures is not appropriate." With the May, 1985, policy statement, the Board reaffirmed this policy. If the Board adopts a maximum transfer limit as a risk reduction measure, the Board will similarly consider it an unacceptable use of Fedwire to avoid the intent of the transfer limit, such as by multiple deliveries at the same time for the account of the same customer, unless the securities were already in position at the time of the order. Reserve Banks would monitor the book entry wire and take appropriate action to end violations of the Board's policy.

As with levels for sender net debit caps, the Board is intentionally

setting the transfer limit at a high level, and plans to reduce the level over time as more experience is gained. The Board is interested in the public's view as to whether the initial limit should be set lower, such as $25 million.

3. Netting Arrangements

The Federal Reserve System will continue to monitor private sector initiatives to develop a non-Federal Reserve facility for netting of securities trades made prior to a given date. The Board understands that participants of such a facility would be mainly dealers and brokers. The facility would net positions multilaterally and then settle the nets through the Federal Reserve's book-entry wire. Such an approach, by reducing daylight exposure and intra-day credit risks, especially at the large clearing banks, could significantly reduce Federal Reserve market exposure. The Board believes this private sector initiative should be monitored closely, however, to ensure that these reduce both Federal Reserve and systemic risks in a fashion that provides adequate safeguards and limitations within the netting system.

The Board proposes that any private network desiring to obtain Federal Reserve net settlement services for the clearing of U.S. Treasury or agency securities would have to provide intra-day

Company Names (DIALOG Generated): Federal Reserve Bank ; Public Securities Association ; Reserve Bank of New York ; Reserve Banks